



Sinha, Abhinav (2014) Molecular basis of gametocytogenesis in malaria parasites. PhD thesis.

<http://theses.gla.ac.uk/5580/>

Copyright and moral rights for this thesis are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

# **Molecular basis of gametocytogenesis in malaria parasites**

Abhinav Sinha MD, MSc, MRes

**Thesis submitted in fulfilment of the requirements for the  
Degree of Doctor of Philosophy**

The School of Life Sciences  
College of Medical, Veterinary and Life Sciences  
University of Glasgow

September 2014

## Abstract

Malaria, a parasitic disease caused by five species of the protozoan parasite *Plasmodium*, still kills an estimated 0.6 million people each year, almost all in the third world African countries. With renewed emphasis on global eradication of malaria, genome-based discovery of novel anti-transmission candidates has been identified as one of the priority research areas for the immediate future.

The aim of this study was to exploit the benefits of a combination of classical forward/reverse genetics approaches, flow cytometry and high throughput whole genome sequencing to examine the molecular basis of gametocytogenesis in the rodent malaria parasite, *P. berghei*. *Plasmodium* is known to spontaneously generate gametocyte non-producing (GNP) mutants if asexually maintained for a long time. Using a sex-specific fluorescently labelled *P. berghei* parental line, ten parallel isogenic lineages were asexually maintained in mice by repeated mechanical passage for a year. Three, out of the potential ten lines, developed the GNP phenotype at the end of the study.

The three GNP and their isogenic parental lines were sequenced using the Illumina platform and the sequence variations analysed. It was discovered that one single gene, a transcription factor with an AP2 domain (PBANKA\_143750), was uniquely mutated in all the three *de novo* GNP mutants and also in the two pre-existing GNP mutants. The gene, called AP2-G, was thus implicated in regulating a switch associated with commitment to gametocytogenesis.

Further conclusive evidence was generated using targeted AP2-G knockout studies (producing the GNP phenotype) and complementation studies in the AP2-G mutants (restoring the WT phenotype). AP2-G was also shown to recognize and bind to a conserved DNA motif in the selected gene promoters in a sequence-specific manner. Inhibition of this interaction by a synthetic customized polyamide compound, ISS-15, was also demonstrated *in vitro*.

Collectively, the work done in the thesis (together with simultaneous independent evidence of involvement of the *P. falciparum* orthologue of AP2-G in gametocytogenesis) established AP2-G as the critical regulator of the commitment to gametocytogenesis in the form of a molecular switch.

# Table of contents

Abstract .....	ii
Table of contents .....	iii
List of tables.....	ix
List of figures.....	x
Acknowledgements .....	xiii
Author's Declaration .....	xvi
Definitions / Abbreviations / Acronyms.....	xvii
 Chapter 1    Introduction .....	 21
1.1 Global burden of malaria .....	22
1.2 Clinical features of malaria .....	23
1.3 Challenges in malaria control .....	24
1.3.1 Artemisinin resistance .....	25
1.3.2 Malaria vaccines .....	25
1.4 Malaria eradication - a distant possibility.....	26
1.5 Future priority areas for malaria research.....	26
1.6 Life cycle of <i>Plasmodium</i> .....	28
1.6.1 Asexual development in vertebrate host.....	29
1.6.1.1 Pre-erythrocytic development .....	29
1.6.1.2 Intra-erythrocytic development .....	31
1.6.1.3 Post-erythrocytic development .....	37
1.6.2 Sexual development .....	39
1.6.2.1 Gametocytogenesis .....	39
1.6.2.2 Gametogenesis, fertilization and ookinete formation .....	42
1.6.2.3 Ookinete to oocyst transformation.....	46
1.6.2.4 Oocyst development and sporozoite differentiation .....	47
1.7 Gametocytogenesis - a sexual developmental switch.....	50
1.7.1 Commitment to gametocytogenesis.....	50
1.7.2 Factors associated with the commitment switch .....	51
1.7.3 Timing of commitment - when does the commitment occur? .....	53
1.7.4 Spontaneous loss of gametocytogenesis and genomic changes .....	54
1.8 Recent technological advances in whole genome analyses .....	57
1.8.1 Sequencing methods .....	57
1.8.2 Next generation sequencing.....	57
1.8.3 Genome assembly .....	59
1.8.4 Structural variants .....	62
1.9 Transcription and its regulation .....	62
1.9.1 General introduction .....	62
1.9.2 Transcription factors .....	63
1.9.3 Epigenetic control of gene expression.....	65
1.9.4 The promoter region of the gene.....	68
1.9.5 Regulation of gene expression in <i>Plasmodium</i> .....	69
1.9.5.1 Introduction .....	69
1.9.5.2 Transcription factors in <i>Plasmodium</i> .....	69
1.9.5.3 Epigenetic regulation of gene expression in <i>Plasmodium</i> .....	71
1.9.5.4 Nuclear structure and gene regulation .....	73
1.9.5.5 Post-transcriptional regulation of gene expression .....	75

1.9.6 ApiAP2 transcriptor factor family in <i>Plasmodium</i> .....	76
1.9.6.1 Introduction .....	76
1.9.6.2 Conservation and functional significance .....	76
1.9.6.3 Structure of ApiAP2 proteins .....	77
1.9.6.4 Structural determinants of ApiAP2-DNA binding .....	78
1.9.6.5 DNA binding specificities.....	79
1.9.6.6 ApiAP2 target genes .....	79
1.9.6.7 Role of ApiAP2 family in <i>Plasmodium</i> gene regulation.....	79
1.9.6.7.1 ApiAP2-O .....	80
1.9.6.7.2 ApiAP2-Sp .....	81
1.9.6.7.3 ApiAP2-L .....	81
1.9.6.7.4 PfSIP2.....	82
1.9.6.8 Combinatorial gene expression and ApiAP2 TFs.....	82
1.9.6.9 Summary and future perspectives.....	83
1.10 <i>Plasmodium berghei</i> - the rodent malaria parasite .....	84
1.10.1 Model malaria parasites .....	84
1.10.2 Genetic engineering techniques in <i>P. berghei</i> .....	86
1.10.2.1 Transfections in <i>P. berghei</i> .....	86
1.10.2.2 Studying essential genes in <i>P. berghei</i> .....	87
1.10.2.3 Gene complementation in <i>P. berghei</i> .....	88
1.10.2.3.1 Complementation techniques.....	89
1.10.2.4 Recent advances in <i>P. berghei</i> genetic engineering techniques ....	91
1.10.2.5 Recombineering based constructs for genetic manipulation of <i>P. berghei</i> .....	92
1.10.3 The genome of <i>P. berghei</i> .....	93
1.10.4 Challenges in assembly of <i>P. berghei</i> genome .....	94
1.11 Aims and objectives .....	95
 Chapter 2     Materials and Methods.....	 96
2.1 General parasitological methods .....	97
2.1.1 Parasite growth <i>in vivo</i> .....	97
2.1.1.1 Animal handling and maintenance .....	97
2.1.1.2 Infecting the mice with parasites .....	97
2.1.1.3 Maintaining infection between mice (weekly mechanical passage) .	99
2.1.1.4 Mosquito handling and maintenance.....	100
2.1.2 Parasite growth <i>in vitro</i> .....	101
2.1.2.1 Schizont culture .....	102
2.1.2.2 Ookinete culture and count .....	102
2.1.2.3 Exflagellation assay .....	103
2.1.2.4 Cryopreservation and reactivation .....	103
2.1.3 Giemsa staining.....	104
2.1.4 Collection of blood from the infected mouse .....	104
2.1.5 Drugs used during parasite growth and maintenance .....	105
2.2 General molecular biology methods .....	106
2.2.1 Isolation of genomic DNA .....	106
2.2.2 Polymerase Chain Reaction (PCR) .....	107
2.2.3 Oligonucleotides.....	108
2.2.4 Southern analysis.....	108
2.2.5 Isolation of total RNA.....	109
2.2.6 Preparation of cDNA and RT-PCR .....	110
2.2.7 Northern analysis.....	111
2.2.8 Generation of gene knockout constructs .....	112

2.2.9 Generating gene complementation constructs .....	112
2.2.10 Parasite transfection .....	113
2.2.10.1 Preparation of DNA for transfection .....	113
2.2.10.2 Preparation of schizonts for transfection .....	116
2.2.11 Cloning of transgenic parasites by limiting dilution .....	116
2.3 Recombinant protein production .....	117
2.4 Electro-mobility Shift Assays (EMSA) .....	119
2.5 Flow cytometry .....	120
2.5.1 Hoechst staining of the samples .....	121
2.5.2 Instrument settings for the MACSQuant® Analyzer .....	121
2.6 Sequencing and bioinformatic methods .....	123
2.6.1 Preparation of DNA for whole genome sequencing .....	124
2.6.2 Illumina/Solexa whole genome sequencing and assembly of the GNP and their WT isogenic parental lines .....	125
2.6.3 Mapping and variant calling .....	126
2.7 Parasite lines used in the study .....	127
2.7.1 For assessment of gametocytogenesis .....	127
2.7.2 For negative control .....	128
2.7.3 Lines being used for reference sequencing .....	129
 Chapter 3     Generation of <i>de novo</i> gametocyte non-producer mutants .....	 130
3.1 Introduction .....	131
3.2 Flow cytometry (FACS) was used to monitor gametocytemia .....	132
3.3 Asexual maintenance of 10 isogenic lines .....	132
3.4 Kinetics of growth among ten isogenic lines .....	133
3.4.1 Kinetics of growth in Line 1 .....	135
3.4.2 Kinetics of growth in Line 2 .....	136
3.4.3 Kinetics of growth in Line 3 .....	137
3.4.4 Kinetics of growth in Line 4 .....	138
3.4.5 Kinetics of growth in Line 5 .....	139
3.4.6 Kinetics of growth in Line 6 .....	140
3.4.7 Kinetics of growth in Line 7 .....	140
3.4.8 Kinetics of growth in Line 8 .....	142
3.4.9 Kinetics of growth in Line 9 .....	143
3.4.10 Kinetics of growth in Line 10 .....	144
3.5 Lines 9, 7 and 8 did not produce gametocytes .....	145
3.5.1 Giemsa-stained peripheral blood smears .....	145
3.5.2 FACS analyses .....	146
3.5.3 Exflagellation assays .....	148
3.5.4 Ookinete cultures .....	148
3.5.5 Transmission through mosquitoes .....	149
3.6 Characterization of the <i>de novo</i> GNP mutants .....	149
3.6.1 Asexual growth advantage .....	149
3.6.2 Sexual growth .....	150
3.6.2.1 Absence of <i>P28</i> transcripts on RT-PCR .....	150
3.6.2.2 Absence of <i>P28</i> transcripts on northern analysis .....	151
3.6.3 Confirmation of presence of the GFP/RFP expression cassette in the <i>de novo</i> GNP mutants .....	152
3.6.3.1 PCR confirmation of the presence of the expression cassette .....	153
3.6.3.2 Evidence from whole genome re-sequencing data .....	154
3.6.4 Selectable marker cassette and <i>de novo</i> GNP mutants .....	155
3.6.4.1 Removal of the SM cassette from the GNP lines .....	156

3.6.4.2 Cloning of the negatively selected GNP lines .....	157
3.7 Evidence from mathematical modelling of the kinetics .....	158
3.7.1 The distribution of the data for the controls .....	158
3.7.2 Sex ratio .....	159
3.7.3 Predicting post-mutational timescale changes in the parasite population within a mouse .....	160
3.7.4 Variance partitioning .....	161
3.8 Summary and discussion .....	163
3.9 Future directions .....	166
 Chapter 4   Whole genome re-sequencing, assembly and genome-wide comparison of de-novo GNP mutants and the isogenic parental lines .....	168
 4.1 Introduction .....	159
4.2 Illumina / Solexa whole genome sequencing and assembly of the GNP and their wild-type isogenic lines .....	170
4.3 Mutations identified in the study lines after mapping on the reference assembly lines .....	171
4.3.1 Mutations within the open reading frames in the GNP study lines after mapping onto the 820 reference assembly .....	172
4.3.2 Mutations within the open reading frames in the GNP study lines after mapping onto the ANKA reference assembly .....	174
4.3.3 Mapping mutations within 1 kb upstream and downstream of the open reading frames in the study lines using the ANKA reference assembly	175
4.3.4 Summary of filtered mutations .....	176
4.4 One gene uniquely mutated in all the GNPs .....	178
4.4.1 Evidence from pre-existing GNP lines .....	180
4.4.1.1 Mutations in PBANKA_011210 .....	181
4.4.1.2 Mutations in PBANKA_141570 .....	181
4.5 Confirmation of PBANKA_143750 mutations by direct sequencing .....	181
4.6 The ApiAP2 family of transcription factors and AP2-G .....	182
4.6.1 AP2-G DBDs are highly conserved across <i>apicomplexa</i> .....	183
4.6.2 Expression profile of PbAP2-G .....	185
4.6.3 Localization studies; GFP- and HA-tagged PbAP2-G .....	186
4.7 Summary and discussion .....	187
4.7.1 Whole-genome sequencing and GNP mutations .....	187
4.7.2 <i>De novo</i> GNP mutations and PBANKA_143750 .....	189
4.7.3 Supplementary evidence from WGS of pre-existing GNP lines .....	191
 Chapter 5   Gene knockout studies confirm the involvement of AP2-G in commitment to gametocytogenesis .....	193
 5.1 Introduction .....	194
5.2 The AP2-G knockout length variants .....	195
5.3 The knockout genotype .....	196
5.3.1 PCR analyses of the knockout mutants .....	197
5.3.1.1 The double-crossover (conventional) full length AP2-G orf knockout mutant .....	197
5.3.1.2 The double-crossover (recombineering) full length AP2-G orf knockout mutant .....	198

5.3.1.3 The double-crossover (conventional) partial length <i>AP2-G</i> orf knockout mutant .....	200
5.3.1.4 The double-crossover (conventional) <i>AP2-G</i> DNA binding domain (DBD) knockout mutant .....	202
5.3.2 Southern analyses of cloned knockout mutants .....	204
5.3.2.1 Cloning of <i>AP2-G</i> knockout variants .....	204
5.3.2.2 Southern analyses of the clones .....	204
5.4 The knockout phenotype .....	206
5.4.1 Giemsa-stained peripheral smears .....	206
5.4.2 Flow cytometry-based evidence .....	207
5.4.3 Exflagellation assays .....	209
5.4.4 Ookinete cultures .....	209
5.4.5 Northern analyses of the mutants .....	210
5.4.6 RT-PCR analyses .....	211
5.4.7 Transmission through mosquitoes .....	213
5.5 Summary and discussion .....	214
 Chapter 6 Gene complementation in gametocyte non-producer mutants to restore their commitment to gametocytogenesis .....	217
6.1 Introduction .....	218
6.2 Complementation strategies to restore gametocytogenesis in the <i>PbAP2-G</i> mutants .....	219
6.2.1 Complementation strategies for natural <i>PbAP2-G</i> mutants .....	219
6.2.1.1 Complementation of natural <i>PbAP2-G</i> GNP mutants with PbG01_COMP-UP construct .....	220
6.2.1.1.1 Phenotypic and genotypic characterization of the PbG01_COMP-UP complemented natural <i>PbAP2-G</i> GNP mutants .....	222
6.2.1.1.2 Possible explanations for the failed complementation in G367 and G368 .....	224
6.2.1.1.3 Genotypic and phenotypic confirmation of G399 and G400 ....	227
6.2.1.1.4 Repairing the promoter in lines G399 and G400 .....	228
6.2.1.1.5 Negative selection of cloned repaired lines to reduce the <i>PbAP2-G</i> promoter interruption .....	230
6.2.1.2 Complementation of natural <i>PbAP2-G</i> GNP mutants with PbG01_COMP-DOWN construct .....	231
6.2.1.2.1 Genotypic and phenotypic characterization of the GNP mutants complemented with PbG01_COMP-DOWN .....	233
6.2.2 Complementation strategies for genetically engineered <i>PbAP2-G</i> knockout mutants .....	238
6.2.2.1 Complementation of the knockouts using full-length functional copy of <i>PbAP2-G</i> .....	240
6.2.2.1.1 Complementation of the full length <i>PbAP2-G</i> knockout .....	241
6.2.2.1.2 Complementation of the <i>PbAP2-G</i> DNA binding domain (DBD) knockout .....	249
6.2.2.1.3 Complementation of the <i>PbAP2-G</i> partial open reading frame knockout .....	253
6.2.2.2 Complementation of the knockouts using mutation-free PCR product .....	255
6.3 Summary and discussion .....	258



Chapter 7	Analysing protein-DNA interactions .....	264
7.1	Introduction .....	265
7.2	GST tagged DNA Binding Domain (DBD).....	266
7.3	AP2-G DBD binds to expected DNA sequence.....	268
7.3.1	The DBD recognizes a DNA sequence/motif: experimental set up .....	268
7.3.2	The DBD binds to the motif identified by PBM analyses in a sequence specific manner .....	270
7.3.3	The binding is relieved by a sequence-specific competitor .....	273
7.4	Polyamides and competition for DNA binding .....	274
7.5	Summary and discussion .....	276
Chapter 8	General discussion .....	278
Appendices.....		290
A1-A10:	Weekly parasitemia & gametocytemia in ten experimental lines ...	291
A11:	Details of mutated gene products temporarily considered under reduced ranking priority.....	301
A12:	Comprehensive list of selected prioritised mutations grouped according to their co-occurrence in the studied GNP lines.....	313
A13:	Summary of all the oligonucleotides used in the study .....	319
A14:	List of chemicals and reagents used .....	325
References .....		330

## List of Tables

Table 1-1. Summary of some known histone modifications in eukaryotes .....	74
Table 2-1. A summary of basic algorithms used in generating a reference assembly, mapping and variant calling .....	127
Table 3-1. Diagnostic PCR to detect the presence of the GFP/RFP expression cassette in the GNP lines.....	153
Table 3-2. Summary of weekly passages of the ten isogenic lines.....	164
Table 4-1. Parameters of WGS of <i>P. berghei</i> GNP and WT lines used.....	170
Table 4-2. A general schematic of the plan for analyses and classification of total mutations identified by comparative mapping.....	172
Table 4-3. Total mutations within the coding sequences as identified by comparative mapping on WT 820 & filters used to refine .....	173
Table 4-4. Total mutations within the coding sequences as identified by comparative mapping on WT line ANKA & filters used to refine.....	174
Table 4-5. Total mutations within the coding sequences & 1 kb-UTRs as identified by comparative mapping on WT line ANKA and filters used to refine them .....	175
Table 4-6. A summary of filtered mutations .....	176
Table 4-7. Final merged list of the number of mutations after filtering for the reduced ranking priority mutations.....	177
Table 4-8. Distribution & characterization of mutations across the <i>AP2-G</i> orf.	190
Table 5-1. PCR for <i>PbAP2-G</i> complete orf knockout integration check.....	197
Table 5-2. PCR for recombineering-based <i>PbAP2-G</i> complete orf knockout 3' integration check .....	199
Table 5-3. PCR for <i>PbAP2-G</i> partial orf knockout integration check .....	201
Table 5-4. PCR for <i>PbAP2-G</i> DBD knockout integration check.....	202
Table 5-5. Southern analyses of the three <i>PbAP2-G</i> cloned knockout variants.	205
Table 5-6. Phenotype of the three <i>PbAP2-G</i> knockout variants.....	210
Table 6-1. PCR for recombineering-based <i>PbAP2-G</i> complete orf complementation integration check in the GNP9 genome using the PbG01-COMP-UP construct .....	220
Table 6-2. PCR for recombineering-based <i>PbAP2-G</i> complete orf complementation integration check in the WT line 820 genome using PbG01-COMP-UP construct .....	225
Table 6-3. PCR for identifying the selectable marker cassette in the 5'UTR of "complemented" WT line 820 genome using PbG01-COMP-UP construct .....	229
Table 6-4. PCR for identifying the selectable marker cassette in the 3'UTR of complemented GNP9 genome using PbG01-COMP-DOWN construct.....	233
Table 6-5. PCR for recombineering-based <i>PbAP2-G</i> complete orf complementation integration (negative ko) check in the G401cl1 genome using PbG01 construct.....	243
Table 6-6. PCR for recombineering-based <i>PbAP2-G</i> complete orf complementation integration check in the G401cl1 genome using the PbG01 construct .....	247
Table 6-7. PCR for recombineering-based <i>PbAP2-G</i> complete orf complementation integration check in the G418cl6cl3 genome using the PbG01 construct .....	250
Table 6-8. Comprehensive summary of complementation strategies used.....	260
Table 7-1. Expected sizes of GST-tagged AP2-G DNA binding domains .....	267

## List of Figures

Figure 1-1. Global malaria ( <i>P. falciparum</i> ) endemicity map .....	22
Figure 1-2. Global malaria endemicity map showing the distribution of the <i>Plasmodium spp.</i> .....	23
Figure 1-3. Global malaria map showing the distribution of countries categorized as being malaria-free, countries eliminating malaria and countries controlling malaria .....	27
Figure 1-4. The life cycle of <i>Plasmodium</i> .....	30
Figure 1-5. Intra-erythrocytic development of <i>P. falciparum</i> .....	34
Figure 1-6. Development of malaria parasite inside the mosquito vector .....	43
Figure 1-7. The technique of bridge amplification used by Illumina/Solexa for generating the paired-end sequencing .....	60
Figure 1-8. The effect of length of repeats on genome assembly .....	60
Figure 2-1. A cartoon of the pL0034 plasmid used for generating the vectors for targeting the <i>AP2-G</i> gene .....	113
Figure 2-2. A flow diagram representing the major pipeline for whole genome sequencing, assembly, variant calling, and analyses of mutations .	124
Figure 2-3. A cartoon of the parental WT line PBANKA 820cl1m1cl1 locus .....	128
Figure 3-1. FACS plots showing gametocytemia in the experimental controls .	133
Figure 3-2. Parasitemia of the serially passaged line 1 at passage .....	135
Figure 3-3. Parasitemia of the serially passaged line 2 at passage .....	136
Figure 3-4. Parasitemia of the serially passaged line 3 at passage .....	137
Figure 3-5. Parasitemia of the serially passaged line 4 at passage .....	138
Figure 3-6. Parasitemia of the serially passaged line 5 at passage .....	139
Figure 3-7. Parasitemia of the serially passaged line 6 at passage .....	140
Figure 3-8. Parasitemia of the serially passaged line 7 at passage .....	141
Figure 3-9. Parasitemia of the serially passaged line 8 at passage .....	142
Figure 3-10. Parasitemia of the serially passaged line 9 at passage .....	143
Figure 3-11. Parasitemia of the serially passaged line 10 at passage .....	144
Figure 3-12. Gametocytemia of the ten serially passaged lines at passage ....	145
Figure 3-13. FACS analyses of the three de novo GNP lines .....	147
Figure 3-14. Multiple bar chart demonstrating the absence of Exflagellation Centres and Ookinetes in the three <i>de novo</i> GNP lines .....	148
Figure 3-15. Development of parasitemia after transmission bite-back experiment in the three de novo GNP lines and a WT control ....	149
Figure 3-16. RT-PCR analyses of <i>ALBA-3</i> and <i>P28</i> from the three <i>de novo</i> GNP lines and a WT control .....	151
Figure 3-17. Northern blot analyses of <i>P28</i> expression in the three <i>de novo</i> GNP lines and a WT control .....	152
Figure 3-18. Schematic of the <i>P. berghei</i> 230p locus on chromosome 3 and PCR showing the presence drug selectable marker cassette in the WT parental line 820 and the three <i>de novo</i> GNP lines .....	153
Figure 3-19. Drug assay showing the presence of drug selectable marker cassette in the WT parental line 820 .....	155
Figure 3-20. Schematic of the <i>P. berghei</i> 230p locus on chromosome 3 before and after drug selectable marker cassette recycling via negative selection.....	157
Figure 3-21. Diagnostic PCR across the <i>P. berghei</i> 230p locus on chromosome 3 before and after drug selectable marker cassette recycling in WT parental line 820 and the three de novo GNP lines .....	158

Figure 3-22. Plotting of the raw data after a logistic transformation in the ten isogenic serially passaged lines .....	159
Figure 3-23. Statistically modelled female and male gametocyte kinetics of the ten isogenic serially passaged lines .....	160
Figure 3-24. Change in gametocytemia kinetics due to the presence of even one parasite harbouring the GNP mutation .....	161
Figure 3-25. Boxplots showing the different possible sources of variability within the experimental setting .....	163
Figure 3-26. Representative growth kinetics from a gametocyte producer and two GNP lines .....	166
Figure 4-1. Cartoon summarizing whole genome sequencing based identification of SNPs on comparative mapping of GNP sequences on the WT 820 parental genome assembly .....	171
Figure 4-2. Breakdown of mutated gene products temporarily considered as reduced ranking priority .....	177
Figure 4-3. Schematic of <i>PbAP2-G</i> open reading frame showing the location of the 3 <i>de novo</i> GNP mutations & 2 pre-existing GNP mutations relative to translation start site of <i>AP2-G</i> & its DBD .....	179
Figure 4-4. Portions of multiple amino-acid sequence alignment from WT parental line 820 and the 3 <i>de novo</i> GNP lines .....	183
Figure 4-5. BLAST search result hits with <i>AP2-G</i> DBD sequence of <i>P. berghei</i> as the query sequence .....	184
Figure 4-6. Expression profiling of <i>PbAP2-G</i> across different developmental stages of synchronised <i>P. berghei</i> infection .....	186
Figure 4-7. Flow diagram of filtration of identified mutations in GNP lines ....	188
Figure 5-1. Schematic of <i>PbAP2-G</i> knockout length variants .....	196
Figure 5-2. Schematic of <i>PbAP2-G</i> conventional DXO complete orf knockout strategy .....	198
Figure 5-3. Schematic of <i>PbAP2-G</i> recombineering-based DXO complete orf knockout strategy .....	199
Figure 5-4. Schematic of <i>PbAP2-G</i> conventional DXO partial orf knockout strategy .....	201
Figure 5-5. Schematic of <i>PbAP2-G</i> conventional DXO DBD knockout strategy ..	203
Figure 5-6. Schematic of Southern analyses of the three <i>PbAP2-G</i> cloned knockout variants .....	205
Figure 5-7. FACS plots for the 3 <i>PbAP2-G</i> knockouts with WT line 820 .....	207
Figure 5-8. Northern blot analyses of the WT and the three <i>PbAP2-G</i> cloned knockout variants .....	211
Figure 5-9. Agarose gel electrophoresis of RT-PCR for <i>ALBA2</i> and <i>P28</i> in the three <i>PbAP2-G</i> knockout variants .....	212
Figure 5-10. Post-transmission parasitemia kinetics in the line G401cl1 .....	214
Figure 6-1. Schematic of complementation strategy used to repair the m9 mutation in the <i>PbAP2-G</i> (m9 mutation repaired) .....	221
Figure 6-2. Schematic of complementation strategy used to repair the m9 mutation in the <i>PbAP2-G</i> (m9 mutation intact) .....	223
Figure 6-3. Multiple sequence alignment of GNP9 repaired and unrepaired clones .....	224
Figure 6-4. Schematic of strategy used to interrupt the WT <i>AP2-G</i> promoter with PbG01-COMP-UP complementation construct .....	226
Figure 6-5. FACS plots showing gametocytemia in the promoter-interrupted lines G399-400 .....	227
Figure 6-6. Negative selection for the promoter-interrupted lines G399-400 ..	228

Figure 6-7. Diagnostic PCR for the detection of SMC in the promoter-interrupted lines G399-400 .....	230
Figure 6-8. FACS plots showing gametocytemia in the negatively selected promoter-interrupted lines - G399m1 and G400m1 .....	231
Figure 6-9. Schematic of the complementation strategy used to repair the naturally acquired GNP mutations .....	232
Figure 6-10. PCR to show the integration in the GNP lines 9 and 7 complemented with PbG01-COMP-DOWN .....	235
Figure 6-11. FACS plots to show gametocytemia in the GNP lines 9 and 7 complemented with PbG01-COMP-DOWN .....	236
Figure 6-12. Giemsa-stained peripheral blood smears to show gametocytes in GNP lines 9 and 7 complemented with PbG01-COMP-DOWN .....	238
Figure 6-13. Schematic of the complementation strategy used to repair the AP2-G knockout length variants .....	240
Figure 6-14. Schematic of the complementation strategy used to repair the <i>PbAP2-G</i> complete orf knockout variant (G401cl1).....	241
Figure 6-15. Negative selection of the repaired <i>PbAP2-G</i> complete orf knockout variant with PbG01 recombineering based construct ...	242
Figure 6-16. Diagnostic PCRs to check integration and WT orf in the G401cl1 line complemented with PbG01 construct .....	243
Figure 6-17. Negative selection of the repaired <i>PbAP2-G</i> complete orf knockout variant with PbG01 recombineering based construct (repeat) ...	246
Figure 6-18. Diagnostic PCRs to check integration and WT orf in the G401cl1 line complemented with PbG01 construct .....	247
Figure 6-19. Negative selection of the repaired <i>PbAP2-G</i> complete orf knockout variant with PbG01 recombineering based construct (third time) .....	249
Figure 6-20. Schematic of the complementation strategy used to repair the DBD knockout variant (G418cl6cl3) .....	250
Figure 6-21. Diagnostic PCRs to check integration & WT orf in the G418cl6cl3 (DBD knockout) line complemented with PbG01 construct .....	251
Figure 6-22. Negative selection of the repaired <i>PbAP2-G</i> DBD knockout variant with PbG01 recombineering based construct .....	252
Figure 6-23. Schematic of the complementation strategy used to repair the partial orf knockout variant (G529cl2) .....	254
Figure 6-24. Negative selection of the repaired <i>PbAP2-G</i> partial length knockout variant with PbG01 recombineering based construct .....	254
Figure 6-25. Schematic of PCR-based complementation strategy for the DBD knockout of <i>PbAP2-G</i> .....	255
Figure 6-26. Negative selection of the transfectants following PCR-based complementation of the <i>PbAP2-G</i> DBD knockout .....	257
Figure 6-27. Summary of various complementation strategies used. ....	262
Figure 7-1. SDS-PAGE of GST-tagged AP2-G induced & uninduced lysates .....	267
Figure 7-2. SDS-PAGE image of GST-tagged AP2-G batch-purified eluates .....	268
Figure 7-3. EMSA showing sequence specific binding of GST-tagged AP2-G DBD to its hypothetical recognition motif .....	271
Figure 7-4. EMSA showing sequence specific binding of GST-tagged AP2-G DBD to its real/naturally occurring recognition motifs.....	272
Figure 7-5. EMSA showing dose-dependent competition between labelled and unlabelled AP2-G targets with the GST-tagged DBD .....	274
Figure 7-6. EMSA showing competitive dose-dependent interference of synthetic polyamide ISS-15 on the AP2-G DBD interaction with its recognition motif .....	276

# Acknowledgements

It would not have been possible to perform the work presented here and write this doctoral thesis without the help and support of the kind people around me who have contributed, directly or indirectly, towards successful completion and compilation of the work. I would like to express sincere gratitude to all mentioned below (and many more) which I might not have expressed or even realized while doing the work.

I would like to express my deep gratitude to my research supervisor, Prof Andy Waters for his patient guidance, enthusiastic encouragement and useful critiques of this work. He has taught me, both consciously and unconsciously, how a good experimental research is planned and done. His selfless time and care were sometimes all that kept me going. The joy and enthusiasm he has for his research was contagious and motivational for me, even during tough times in the PhD pursuit.

Specially and most importantly, I wish to acknowledge the generous support provided by Dr Manuel Llinas group (for kindly providing GST-tagged *P. berghei* and *P. falciparum* AP2-G DBD and troubleshooting EMSAs), Dr Oliver Billker group (for kindly supplying us with the recombineering based constructs for gene knockout and complementation and the related troubleshooting), Dr Ishwar Singh (for providing the polyamides ISS-15 and ISS-33), and Dr Jayde Gwathorne (for supplying eLOV-tagged *P. berghei* AP2-G DBD).

Exceptional thanks are due to Dr Thomas D Otto and Ms Mandy Sanders (for performing whole genome sequencing, comparative mapping and analyses for the GNP lines), Dr Nicholas Dickens (for round the clock bioinformatics analyses and support, when and where needed) and Dr Matt Denwood (for performing mathematical modelling on the data).

I am particularly grateful for the advice given by Drs April Williams, Bjorn Kafsack, and Manuel Llinas (all from Princeton University, USA), Drs Claudia Pfander, Ellen Bushell, Katarzyna Modrzynska, Matthew Berriman, and Oliver Billker (all from Wellcome Trust Sanger Institute, Hinxton, UK), Profs Brian Shiels, Daniel Haydon, and Ms Marta Piezko (University of Glasgow, UK) and Dr Glenn Burley (University of Strathclyde, UK) which particularly proved valuable to fine tune my work.

I sincerely thank my assessors Dr Lisa Ranford-Cartwright and Prof Sylke Muller, for their overall guidance, critical overview of the research and suggesting how to keep focused. They both deserve huge credit for keeping me and my progress on track and timely finishing the work.

I would like to express my very great appreciation to all the current and former members of Prof Andy Waters Lab, namely Agnieszka Religa (my spiritual friend-

cum-advisor and for disseminating positive energies all around), Angela McBride (for friendly insectary support), Anne Graham (for directly or indirectly imparting her excellent management skills and for huge technical support whenever my project suffered a setback), Anubhav Srivastava (for extremely friendly advice, continuous on-lab and off-lab support whenever I needed and most importantly, always with a smile), Harshal Patil (for brotherly support, continuous encouragement, and day-to-day technical advice), Katie Hughes (for extremely valuable technical tips that always worked and for providing detailed insights into the work), Luke Starnes (for friendly encouragement and support), Mhairi Stewart (for always helping me when technically stuck specially with EMSA and RNA apart from keeping my working environment light), Nisha Philip (for troubleshooting everything I did wrong and for making me learn how to utilize every second in the lab), Rachael Orr (the wonder lady who had solutions to every problem I faced during my PhD and for always standing by my side to ensure the smooth and uninterrupted running of the project), and Sonya Taylor (for her round the clock support and motivation specially for a lab novice like me).

Specially, I thank the UK Foreign and Commonwealth Office, the Chevening Scholarship and the Wellcome Trust for funding my studies at Glasgow. Without the Trust's overwhelming support, I cannot imagine this work could have been completed almost in time. I would like to offer my special thanks to Prof Bill Cushley, Prof Darren Monckton, Dr Olwyn Byron and Ms Linda Atkinson for providing me with all the support I needed from the Wellcome Trust during my hardship.

I also thank the University of Glasgow for providing me a space and opportunity to learn and advance the scientific knowledge. I specially thank Mulu Gedle for continuous technical and office support.

Assistance provided by the Animal House staff (especially Collin and Margaret), Flow cytometry facility (especially Diane Vaughan and Alasdair Fraser) and Jim Scott was enormous and is greatly appreciated.

I thank the members of the level 5 and 6, Glasgow Biomedical Research Centre, and my Wellcome Trust course mates for friendly support, technical discussions and tips. I particularly thank Abdulsalam Alkhaldi, Cora Jean Meaden, Daniel Tagoe, David Wildridge, Dhilia Lamasudin, Eduard Kerkhoven, Federica Giovani, Georgina Humphreys, Ibrahim Teka, Isabel Vincent, Isabelle Dietrich, Jonathan Mwangi, Joseph Mugasa, Katherina Johnston, Larissa Laine, Lenka Richterova, Lyndsey Plenderleith, Noushin Enami, Olumbe Ajibola, Roy Mwenechanya, and Sultan Alghamdi, to name a few.

I extend sincere thanks to MWG Operon (Germany) for providing me with sequencing results and Fermentas Life Sciences (Canada) for supplying some free samples.

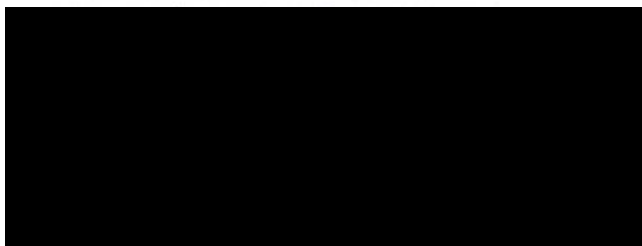
Thanks are also due to Dr Mahadik, the Medical Director of R D Gardi Medical College, Ujjain, India for providing me with all the support and cooperation in writing up this work. A big thanks to Drs Arpit, Deepika, Rashmi, Shikha, Suchita, Tarique and Vibha for proof-reading this piece of work and suggesting appropriate corrections.

To the end, I would like to thank my wife, Parul and sons, Ashar and Zenish for their unspoken support and good wishes. For any errors or inadequacies that may still remain in this work, of course, the responsibility is entirely my own.



## Author's Declaration

I declare that the results presented in this thesis are my own work, except where stated otherwise, and that this work has not been submitted for a degree at any other institution.



Abhinav Sinha

## Definitions/Abbreviations/Acronyms

µg	microgram
µl	microliter
µM	micro-molar
3' UTR	3-prime untranslated region
5-FC	5-Fuorocytosine
5-FCTP	5-Fuorocytosine triphosphate
aa	amino acid
ABACAS	Algorithm-Based Automatic Contiguation of Assembled Sequences
ACT	Artemisinin-based Combination Therapy
ALBA	Acetylation Lowers Binding Affinity
AMA	Apical Membrane Antigen
Amp	Ampicillin
AP2	Apetala 2
AP2/ERF	Apetala 2/Ethylene response factor
AP2-G	Apetala 2 - Gametocytes
AP2-G2	Apetala 2 - Gametocytes-2
AP2-L	Apetala 2 - Liver stage
AP2-O	Apetala 2 - Ookinetes
AP2-Sp	Apetala 2 - Sporozoites
Api AP2	Apicomplexan AP2
AT rich	Adenosine and Thymine rich
ATG	Start codon (Methionine)
BAC	Bacterial Artificial Chromosome
BamHI	<i>Bacillus amyloliquefaciens</i> derived restriction endonuclease
BC	Before Christ
BIR	<i>Berghei</i> Interspersed Repeats
bp	base pairs
BSA	Bovine Serum Albumin
BSG	Basigin
Cam	Chloramphenicol
CDC	Centres for Disease Control
Cdc2	cell division cycle protein 2
cDNA	Complementary DNA
CDPK	Calcium-dependent Protein Kinase
cds	coding sequence
CF11	Fibrous cellulose powder
CFP	Cyan Fluorescent Protein
cGMP	Cyclic Guanosine Mono-phosphate
Cis	regions of non-coding DNA that regulate the transcription of nearby genes
CNV	Copy Number Variation
CO <sub>2</sub>	Carbon dioxide
CS	circumsporozoite
CTRAP	Circumsporozoite- and Thrombospondin Related Anonymous Protein
CTRP	Circumsporozoite- and TRAP Related Protein
d1	day 1
DALY	Disability Adjusted Life Years

dATP	Deoxyadenosine triphosphate
DBD	DNA Binding Domain
DCO/DXO	Double Crossover event
DEPC	Diethylpyrocarbonate
DNA	Deoxyribonucleic acid
DNAse	Deoxyribonuclease
dNTP	Deoxyribonucleotide
DTT	Dithiothreitol
ds	Double-stranded
EcoRI	<i>E. coli</i> derived restriction endonuclease
ECs	Exflagellation Centres
EDTA	Ethylenediaminetetraacetic acid
eGFP	Enhanced Green Fluorescent Protein
EMSA	Electrophoretic Mobility Shift Assay
ERF2	Ethylene Response Factor
et al	et alii (and others)
ext. time	Extension time
FACS	Fluorescence Activated Cell Sorting
FITC	Fluorescein isothiocyanate
Flp/FRT	Flippase/Flippase Responsive Target
FSC/SSC	Forward-scatter and Side-scatter (FACS)
GAF	Gamete Activation Factor
GAP	Glideosome Associated Protein
GC	Guanylyl Cyclase
gDNA	Genomic DNA
GFP	Green Fluorescent Protein
GIMO	Gene-In Marker-Out
GNP	Gametocyte Non-Producer
GPI	Glycosylphosphatidylinositol
GST	Glutathione S-transferase
GW	GateWay
HAT	Histone Acetyl-Transferase
HCl	Hydrochloric acid
HDAC	Histone deacetylases
hdhfr	Human dihydrofolate reductase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HF	High-Fidelity
HindIII	<i>Haemophilus influenzae</i> derived restriction endonuclease
HMG	High Mobility Group
HMT	Histone Methyl Transferase
HP	High (gametocyte) Producer
HSP	Heat Shock Protein
HSP70	Heat Shock Protein 70
HT	Host Targeting
i.p.	Intra-peritoneal
i.v.	Intra-venous
iCORN	Iterative Correction of Reference Nucleotide
IGM	Insertional Gametocyte-deficient Mutants
IMAGE	Iterative Mapping and Assembly for Gap Elimination
IMC	Inner Membrane Complex
IND	Indels
Indels	Insertions/deletions
IPTG	Isopropyl B-D-1-thiogalactopyranoside

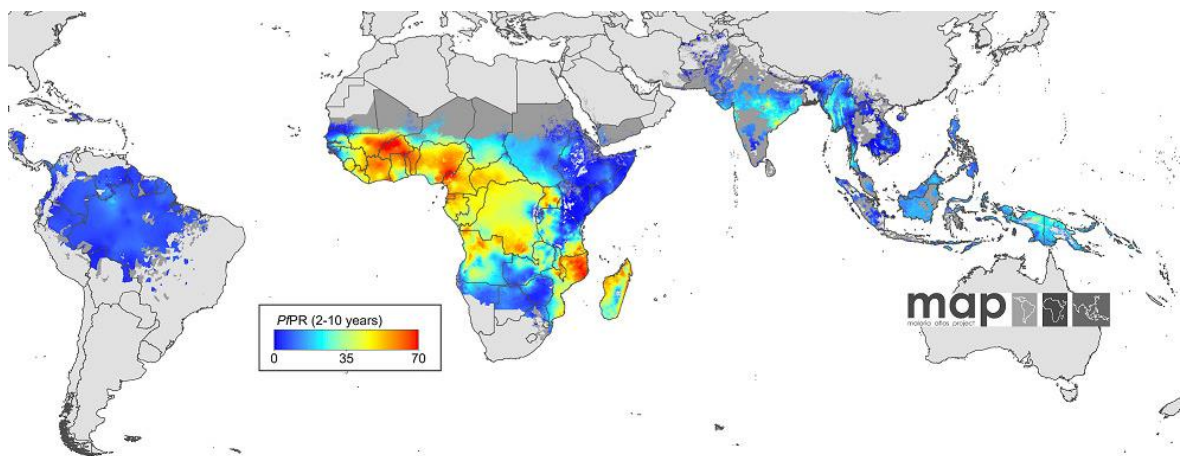
ISS-15	Polyamide mimicking AP2-G DBD
ISS-33	Polyamide mimicking AP2-O DBD
IU	International Units
KDa	Kilo Daltons
ko	Knockout
LB	Luria-Bertani broth
LCCL	Protein domain including Limulus factor C, vertebrate cochlear protein cochlin or coch-5b2 and mammalian late gestation lung protein Lgl1
MAEBL	Apical Membrane Antigen/Erythrocyte Binding-Like protein
MAOP	Membrane Attack Ookinete Protein
MAP	Mitogen Activated Protein
miRNA	Micro-RNA
mM	Milli-molar
MNAse	Monococcal Nuclease
MOPS	3-(N-morpholino) propanesulfonic acid
MSP	Merozoite Surface Protein
MTIP	Myosin A Tail domain Interacting Protein
MyoA	Myosin A
NA	Not applicable
NaAc	Sodium Acetate
NaCl	Sodium Chloride
NaHCO <sub>3</sub>	Sodium bicarbonate
NcoI	<i>Nocardia coralline</i> derived restriction endonuclease
NEB	New England BioLabs
NEK	NIMA-related kinase
NGS	Next Generation Sequencing
NIMA	Never In Mitosis/ <i>Aspergillus</i>
nM	Nano-molar
NotI	<i>Nocardia otitidis</i> derived restriction endonuclease
NS	Negative Selection
OD	Optical Density
ORF	Open Reading Frame
pmol	pico-molar
PAGE	Poly-Acrylamide Gel Electrophoresis
Pb	<i>Plasmodium berghei</i>
PBANKA	Pb ANKA
PbAP2-G	<i>Plasmodium berghei</i> Apetala2 protein for Gametocytogenesis
PBM	Protein Binding Microarray
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PCRMP	<i>Plasmodium</i> cysteine repeat modular proteins
PDE	Phosphodiesterases
PE-A	Phycoerythrin-A channel
PERL	Practical Extraction and Reporting Language
PEXEL	<i>Plasmodium</i> export element
Pf	<i>Plasmodium falciparum</i>
Pfgeco	gametocyte erythrocyte cytosolic protein
PfPKG	<i>Pf</i> cGMP dependent protein kinase
PfPR	<i>Plasmodium falciparum</i> parasite rate
PMSF	Phenylmethanesulfonyl fluoride
PNEP	PEXEL/VTS Negative Export Proteins
PTEX	<i>Plasmodium</i> Translocon of Exported Proteins

Py-Im	Pyrrole-imidazole
RATT	Rapid Annotation Transfer Tool
RBCs	Red Blood Cells
RE	Restriction endonucleases
RFP	Red Fluorescent Protein
RMgm	Rodent Malaria genetically modified database
RNA	Ribonucleic acid
RNAi	RNA interference
rpm	revolutions per minute
RPMI	Roswell Park Memorial Institute medium
rRNA	Ribosomal RNA
RT reaction	Reverse Transcriptase reaction
RT-PCR	Reverse Transcriptase PCR
SAM tools	Sequence Alignment Map tools
SCO	Single Crossover event
SDS	Sodium dodecyl sulphate
SERA	Serine Repeat Antigen
SMALT	Sequence Mapping and Alignment Tool
SMC	Selectable Marker Cassette
SNP	Single Nucleotide Polymorphism
SOC	Super Optimal Broth with catabolite repression
SpeI	<i>Sphaerotilus natans</i> derived restriction endonuclease
SPM-1	Subpellicular microtubular protein - 1
SRS	Short Read Sequences
SSAHA	Sequence Search and Alignment by Hashing Algorithm
SSC	Saline-sodium citrate buffer
SSH	Suppressive Subtractive Hybridization
SUB2	Subtilisin2
SV	Structural Variant
Ta	Annealing temperature
Taq	<i>Thermus aquaticus</i> (polymerase)
TBB	Transmission Bite-Back
TF	Transcription factors
TFBS	TF Binding Sites
Tm	Melting temperature
TO	Theiler's Original (mouse)
TPG	Tags Per Gene
TPK	Tags Per Kb
TRAP	Thrombospondin Related Anonymous Protein
TRX2	Thioredoxin-2
TSS	Transcription start site
UTR	Untranslated regions
UV	Ultraviolet
v/v	volume-by-volume
VLP	Virus Like Particle
VTs	Vacuolar Transport Signal
WGS	Whole Genome Sequencing
WHO	World Health Organization
WT	Wild Type
WTsI	Wellcome Trust Sanger Institute
XA	Xanthiureic acid
yFCU	Bifunctional protein that combines yeast cytosine deaminase and uridyl phosphoribosyl transferase (UPRT)

# **1 General Introduction**

## 1.1 Global burden of Malaria

It has been more than four thousand five hundred years since the first ever recorded historical note of the symptomatology akin to that of Malaria was described in the *Nei Ching* (The Canon of Medicine, ancient Chinese medical writings) back around 2,700 BC (History of Malaria, Centers for Disease Control and Prevention, CDC, Atlanta). Indeed, Malaria is almost certainly one of the ten most ancient diseases of mankind but unfortunately is also the most prevalent parasitic infection (Figure 1-1) and one of the most important infections in humans (Haldar et al., 2007; Clements et al., 2013).



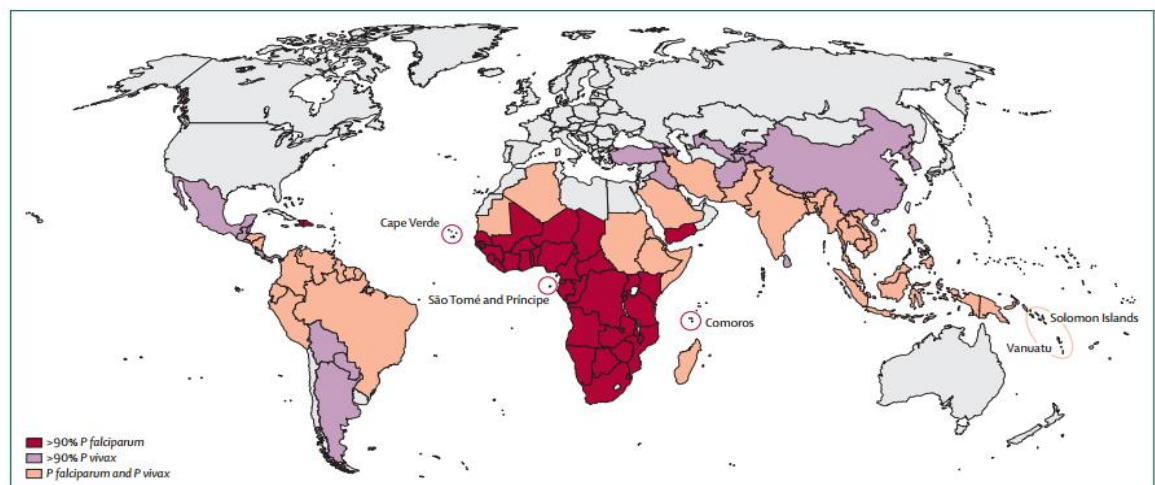
**Figure 1-1. Global malaria (*P. falciparum*) endemicity map showing the regions around the globe with varying degree of age-standardized *P. falciparum* parasite rate ( $PfPR_{2-10}$ )** which describes the estimated proportion of 2-10 year olds in the general population that are infected with *P. falciparum* at any one time, averaged over the 12 months of 2010. Blue and red regions show the extremes of  $PfPR_{2-10}$  (with dark blue signifying 0% and red signifying 70%  $PfPR$ ). Clearly visible are the parts of Africa which still include reddish areas on the map. Source: Gething et al., 2011 – reproduced with permission.

The start of coordinated efforts to fight this disease in the form of health programs is relatively new, dating back to the 1940s (Yekutieli, 1980). Despite our frantic efforts to conquer this scourge, malaria is still annoyingly one of the major killers in the third world - with an estimated 207 million cases globally and 0.62 million or more deaths worldwide in 2012 of which 80% of cases and 90% of deaths were in Sub-Saharan Africa region (Murray et al., 2012; World Malaria Report, 2013). The vast majority (77%) of deaths due to malaria involve children under-five years of age (World Malaria Report, 2013). Not only does it

kill but malaria is responsible for over 219 million “attacks” leading to 800 million days of illness annually in Africa alone (Bremner et al., 2004). Out of the estimated 207 million malaria cases in 2012, ~80% were in the African region alone, followed by 13% in South-East Asia and 6% in the Eastern Mediterranean region (World Malaria Report, 2013). A total of 104 countries of the globe are affected by Malaria putting 3.4 billion people at risk and contributing to 35.4 million DALYs in Sub-Saharan Africa (Global Malaria Action Plan, 2008; World Malaria Report, 2013).

## 1.2 Clinical features of Malaria

Human malaria is caused by five species of *Plasmodium* parasites - *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium ovale*, *Plasmodium malariae*, and most recently included *Plasmodium knowlesi* (Cox-Singh et al., 2008), particularly in South-East Asia (Antinori et al., 2013). *P. falciparum* is responsible for almost all cases of malaria (99%) in African Region (Figure 1-2) whereas it contributes to only 53% cases in South-East Asian region (World Malaria Report, 2013). *P. vivax* causes ~50% of the global malaria burden outside Africa, particularly in the Americas (65%) and South East Asia (47%) (Price et al., 2007; Guerra et al., 2010; World Malaria Report, 2013).



**Figure 1-2. Global malaria endemicity map showing the distribution of the *Plasmodium* spp. (*P. falciparum* and *P. vivax*) across various regions around the globe.** Dark brown regions show a high proportion (>90%) of *P. falciparum* cases whereas the violet areas show predominantly *P. vivax* malaria. Light brown regions have mixed infections with *Plasmodium* species. Source: Feachem, et al., 2010 – reproduced with permission.



The pathogenesis of *P. falciparum* malaria is complex and the clinical spectrum ranges from an asymptomatic infection to mild disease to a rapidly evolving life-threatening disease, severe malaria, involving multiple organ systems and ultimately death. Severe malaria is a clinico-pathological syndrome which may comprise hyperparasitemia ( $>100,000$  per  $\text{mm}^3$ ), cerebral malaria, severe anemia and respiratory distress syndrome (WHO, 2000; Weatherall et al., 2002). Although rarely fatal, *P. vivax* malaria in contrast contributes to a debilitating illness which potentially ends up causing reduced learning and hampered economic growth and development (Wellems et al., 2009). Also, *P. vivax* can remain latent in the liver as hypnozoites which may get activated and result in relapses of malaria later. However, frequent under-reporting of *P. vivax* associated disease and mortality is not uncommon, particularly in children (Poespoprodjo et al., 2009).

### 1.3 Challenges in Malaria control

Despite an estimated 1.1 million malaria deaths averted during the last decade, primarily due to a scale-up of malaria interventions, malaria - an entirely preventable and treatable disease - still, on an average, takes the life of an African child every minute (World Malaria Report, 2012). The problems of severity (especially with *Plasmodium falciparum* infections) and the complex nature of the disease are further exacerbated by a continuing and evolving parasite and vector resistance to antimalarial drugs and insecticides, respectively, which remains one of the most formidable challenges (Tanner and de Savigny, 2008; Campbell, 2009; Mendis et al., 2009; Breman, 2012; Cheeseman et al., 2012; Mideo et al., 2013; Rosenthal, 2013).

A plateauing of the scaled increase in getting access to the most crucial anti-malaria interventions like insecticidal nets, indoor residual spraying, diagnostic testing, and artemisinin-based combination therapies was observed for the very first time in 2010-11. Resistance to artemisinins, the key compounds in artemisinin-based combination therapies or ACTs, has been detected in four countries of South-East Asia (Greater Mekong sub-region: Cambodia, Myanmar, Thailand and Viet Nam), while mosquito resistance to insecticides has been found in 64 countries around the world. Despite the observed changes in parasite

sensitivity to artemisinins, ACTs continue to cure patients provided that the partner drug is still efficacious (World Malaria Report, 2012; Miotto et al., 2013).

### **1.3.1 Artemisinin resistance**

Artemisinin, the wonder drug discovered by Youyou Tu and colleagues in 1981, has been the mainstay for malaria treatment since late 1990's (Miller and Su, 2011). A resistant polymorphic locus conferring slow parasite clearance rates on Artemisinin Combination Therapy (ACT) and subsequently found associated with ACT resistance has been located on *P. falciparum* chromosome 13 thus providing a genetic basis for the phenotype (Cheeseman et al., 2012). These findings have recently been substantiated with the discovery of a molecular marker of artemisinin resistance, a mutant kelch propeller domain protein (termed K13-propeller; PF3D7\_1343700), which was associated with artemisinin resistance, both *in vivo* and *in vitro* (Ariey et al., 2014). Early identification of emerging or low level resistance to artemisinin will be a key target in the immediate future to restrict the resistant parasites and thus controlling the spread of resistance (World Malaria Report, 2013). High throughput methods including deep sequencing data have already been exploited to predict the early development of resistant parasites using differential phenotypic “signatures” of the resistant parasites in a mixed population within a patient (Mideo et al., 2013).

### **1.3.2 Malaria Vaccines**

Malaria candidate vaccines have started to move from laboratories to the field but an effective vaccine still remains a distant dream. With no licensed malaria vaccine yet, the scientific community is eagerly awaiting the results of the Phase 3 Clinical Trial of the candidate vaccine - RTS,S/AS01 to expand its armamentarium against malaria. Although there are approximately 20 other candidates in the pipeline (Clinical Trials), they all are at least 5-10 years behind RTS,S/AS01 (World Malaria Report, 2013). The RTS,S/AS01 consists of a fusion protein involving the carboxy terminus (aa 207-395) of the *P. falciparum* circumsporozoite (CS) antigen (Stoute et al., 1997) and hepatitis B surface antigen virus-like particle (VLP) lyophilized with GlaxoSmithKline's AS01

adjuvant (Reed et al., 2009). Recent Phase 3 clinical trial results estimate statistically significant vaccine efficacy after dose 3 for the age group of 5-17 months (46%) and 6-14 weeks (27%) against clinical malaria (Agnandji et al., 2011; Agnandji et al., 2012; Moorthy et al., 2013a; Otiene C, 2013). With the much needed “second generation” malaria vaccine, probably a transmission blocking one, already on the roadmap, high quality and multi-disciplinary research centered on gametocytogenesis is the need of the hour (Moorthy et al., 2013a & b; WHO “Immunization, Vaccines and Biologicals, 2013).

## **1.4 Malaria eradication – a distant possibility?**

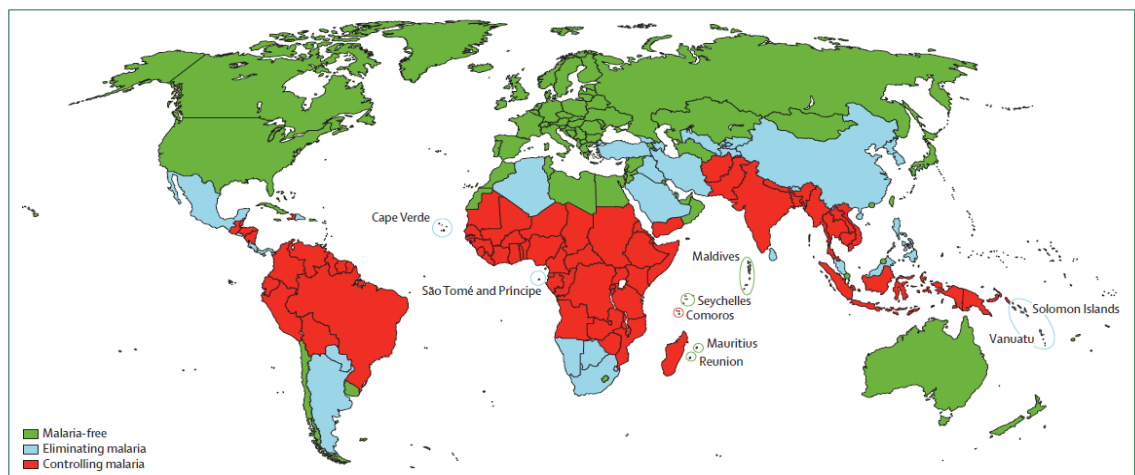
Although in 2007, the Bill and Melinda Gates Foundation, one of the major funders of all global Malaria related activities, reiterated and revived the malaria fraternity to embrace the audacious goal of malaria eradication (Sherrer, 2012), it stands out as a challenge (and perhaps a debate) to the malaria research community worldwide to reconsider the feasibility of malaria eradication in the future (Greenwood, 2009). Visualizing and realizing global malaria eradication as a distant possibility, a tri-tiered strategy towards eradication (Figure 1-3) was developed and endorsed to move on the road to eradication - aggressive malaria control in countries with highest disease burden and death, progressive elimination from endemic margins inwards thus shrinking the malaria map, and intensive research into development of novel interventions and technologies that could be utilized to achieve eradication (Roll back malaria, 2008; Feachem and Sabot, 2008; Feachem et al., 2009).

Scrutinizing the malaria eradication efforts in the past along with currently available interventions and in-depth analysis of parasite biology, it can be easily inferred that approaches directed towards blockage of malaria transmission are potentially the need of the hour (Sinden, 2009) and at the very least a promising approach.

## **1.5 Future priority areas for Malaria research**

It's been a decade since the *Plasmodium* and *Anopheles* genomes were sequenced (Holt et al., 2002; Gardner et al., 2003) but nothing substantial, in

terms of the immediate utility, has yet been translated out of the genome projects (Greenwood and Owusu-Agyeyi, 2012). A huge ray of renewed hope has been levied on the application of the genomic database to find a possible working solution to malaria transmission in the post-genomic period - unless there is a new major technological (probably genomics-based) breakthrough addressing malaria transmission, eradication will always be a distant theoretical goal (The malERA Consultative Group on Basic Science and Enabling Technologies, 2011; Greenwood and Owusu-Agyeyi, 2012; World Malaria Report, 2012). Realisation of immediate insufficiency of the existing tools necessary to achieve the renewed and rejuvenated goal of global malaria eradication in the near future, opens a whole new arena of discovering novel anti-malaria measures but at the same time, puts an enormous pressure on the research community and stake-holders to provide the rays of hope to achieve this mammoth task (Alonso et al., 2011).



**Figure 1-3. Global malaria map showing the distribution of countries categorized as being malaria free (green), countries eliminating malaria (blue) and countries controlling malaria (red).** It is quite apparent that most of the countries where the endemicity is high together with predominant *P. falciparum* infection are still in the malaria control phase. Source: Feachem, et al., (2010) – reproduced with permission.

Considering the limitations of vector control and vaccines, appropriate use of antimalarial drugs remains a cornerstone of malaria control with no solid substitute for ACTs in the near future. Other priority research areas are mechanisms of action of artemisinins, development of new partner drugs with

different modes of actions, and identification of new classes of antimalarial drugs. The best way to stop drug resistance in *P. falciparum* malaria is to eliminate transmission. The biggest challenge in malaria eradication today, along with vaccine development, is how best to marshal an assault on gametocytes with existing measures. In view of the stakes in the fight against emerging ACT resistance, and the goal of worldwide eradication of malaria, gametocytocidal drugs should be used as an additional measure in areas at risk, and development of newer and better drugs, and even vaccines against gametocytes, should be emphasised (Breman, 2012).

With the constant loss of standard anti-malarials to ongoing resistance and no available vaccine for public use in sight, a multi-pronged approach to research, including better understanding of the pathogenesis of the disease and control of severity of the disease is imperative.

## 1.6 Life cycle of *Plasmodium*

Malaria parasites have accompanied us throughout the Darwinian descent and our understanding of the biology of *Plasmodium* is relatively naïve. Any attempt to eliminate (or even control) the disease will necessitate delving deeper into the life cycle and biology of transmission of *Plasmodium* and understand the bottlenecks thereof. There are many missing links that need to be comprehended and sensibly interpreted to solve this puzzling problem and it is this deficiency in understanding of parasite biology that is one of the major obstacles in finding a permanent solution for Malaria (Balu and Adams, 2007). Although ultimately the disease must be understood in humans, much of our current understanding of the pathogenesis of malaria comes from animal models and *in vitro* cultures of *P. falciparum*.

*Plasmodium* has a complex life cycle that switches between an asexual phase in the vertebrate host and a sexual phase followed by asexual phase in the mosquito (Figure 1-4).

### **1.6.1      *Asexual development in the vertebrate host***

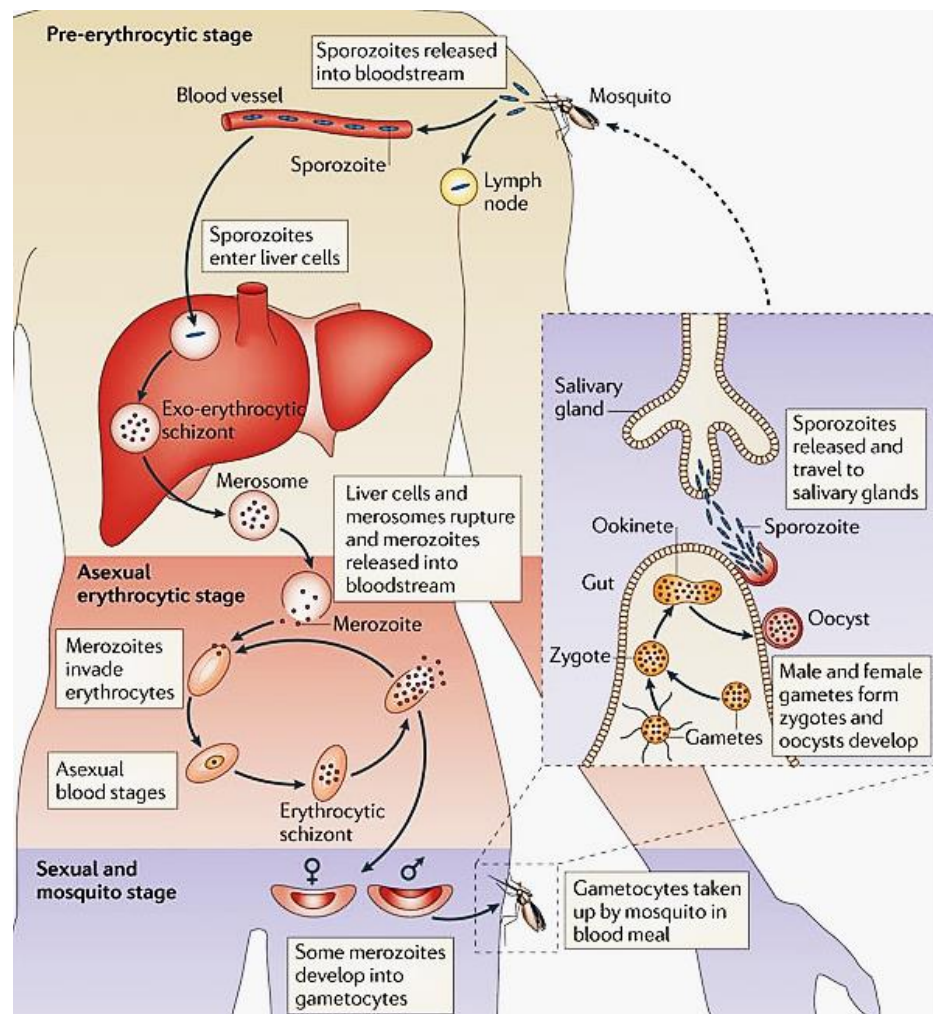
The asexual developmental phase inside the vertebrate host can be divided into a pre-erythrocytic phase (including the development of malaria parasites from the point of entry into the host and development inside the liver up to the entry of parasites into the RBCs of the vertebrate host), an intra-erythrocytic phase (including the development inside the RBCs), and a post-erythrocytic development (between exit from the infected RBCs either to become gametocytes or to re-invade the naïve RBCs) (Figure 1-4).

#### **1.6.1.1      Pre-erythrocytic development**

Haploid sporozoites are inoculated into the skin of a vertebrate host with the bite of an infected female anopheline mosquito. An average inoculum size of 100-125 sporozoites has been shown in studies on rodent malaria parasites in laboratories (Medica and Sinnis, 2005) and the field (Rosenberg et al., 1990; Beier et al., 1991). Recent studies in rodent malaria have demonstrated that there exists a skin sub-cutaneous stage in the course of malaria infection (Ejigiri and Sinnis, 2009) wherein the sporozoites are mostly inoculated in the dermis (Sidjanski and Vanderberg, 1997; Matsuoka et al., 2002). Once injected, sporozoites move along random paths and take between one and three hours to leave the injection site (Yamauchi et al., 2007) and eventually invade blood or lymph vessels (Amino et al., 2006). Not all sporozoites end up in blood vessels - some are destroyed in the skin while some (0.5-5%) may remain at the inoculation site and develop as exoerythrocytic stages (Gueirard et al, 2010; Voza et al, 2012). About 20% migrate to the draining lymph node (Amino et al., 2006; Yamauchi et al., 2007) where the dendritic cells prime CD8(+) T lymphocytes, the first immune response to the parasite (Chakravarty et al., 2007). However, the skin exoerythrocytic stage has only been demonstrated in the rodent malaria parasites and it is an important area to explore whether such stages also occur in primate or human malaria species (Sinnis and Zavala, 2012).

Sporozoites that enter blood vessels move along the endothelium of blood vessels and travel to the liver and cross several hepatocytes before infecting a final one and settling there (Frevert et al., 2005) in a process called productive

invasion (Prudencio et al., 2011). Productive hepatocyte invasion is mediated by invagination of the host cell plasma membrane to form a parasitophorous vacuole which surrounds the invading sporozoites and in which they develop into the next infective stage (Mota et al., 2001). Within the hepatocyte, the sporozoite develops within 47-52 hours via the trophozoite stage into the mature schizont that can contain 1,500-8,000 merozoites (exo-erythrocytic schizogony).



**Figure 1-4. The life cycle of *Plasmodium* showing the different growth stages of the parasite in a vertebrate host and in mosquito vector.** The mosquito life stages, particularly, the development of ookinete and midgut invasion, constitute a major population bottleneck in the parasite life cycle. Source: Sauerwein et al., (2011) – reproduced with permission from Nature.

These merozoites bud off from the infected hepatocyte in membrane bound vesicles termed merozoites after inducing death and detachment of the host hepatocyte. The merozoites squeeze out of the liver through the endothelial cell layer and are thus released into the neighbouring sinusoids (Sturm et al., 2006;

Falae et al., 2010). The merozoites ultimately break off releasing the merozoites into the blood stream where they invade red blood cells and start the blood stage development (Vaughan et al., 2008). The merozoites are invasive and within a period of 30 minutes of their release, they invade a fresh RBC to initiate the erythrocytic cycle.

The liver stage, despite being asymptomatic, is again an important phase of the parasite life wherein the parasite could be attacked to interrupt the cycle at the initial stages of establishment of infection (Prudencio et al., 2011; March et al., 2013). Many of the intriguing immunological cross-talks and other host-parasite interactions (Silvie et al., 2008; Tarun et al., 2008) have been shown to occur during this relatively under-explored stage of the parasite life cycle including the promotion of liver infection in *P. berghei* by host heme-oxygenase 1 (Epiphany et al., 2008). The host cell liver stage transcriptome analyses in *P. berghei* revealed that the host cell transcriptome is modified by the parasitic infection in a timely fashion to best fit the parasites' changing developmental needs (Albuquerque et al., 2009).

### 1.6.1.2 Intra-erythrocytic development

An intracellular life style has been adopted by apicomplexans to, in part, evade the host immune attacks (Dowse and Soldati, 2004; Foller et al., 2009). Intra-erythrocytic development ensures initial rapid expansion and sustained maintenance of the parasite population in the vertebrate host. Mature red cells which are terminally differentiated cells that lack standard biosynthetic pathways and intracellular organelles, offer certain advantages for the parasite.

Since RBCs do not display antigens related to the major histocompatibility complexes on their surfaces, they are a preferred site for immunological escape (Silvie et al., 2008). Also, the abundance of a rich source of nutrients readily available to the parasite, rapid turnover of the red cells, and easy access to the mosquito vector are among other factors behind the specific predilection of *Plasmodium* to red blood cells (Pasvol and Wilson, 1982). However, the absence of standard endocytic and secretory pathways presents a potential hurdle for the fast-growing obligate intracellular parasite. This is partly compensated by a



considerable export of a range of remodeling proteins and virulence factors into the red cell cytoplasm.

This export is mediated through a recently identified specific targeting sequence, termed *Plasmodium* export element (PEXEL) or host targeting (HT) signal (Chang et al., 2008) or vacuolar transport signal (VTS) and a newly discovered protein export machine, *Plasmodium* translocon of exported proteins or PTEX (de Koning-Ward et al., 2009). This specific targeting sequence for translocation is a short consensus sequence RxLxE/Q/D (Hiller et al., 2004; Marti et al., 2004). However, PEXEL/VTS-negative exported proteins (PNEP) have also been identified which suggests a redundancy in the protein export machinery (Grüning et al., 2012; Heiber et al., 2013). Once secreted into the parasitophorous vacuole lumen, the exported proteins are shown to be associated with a putative *Plasmodium* translocon of exported protein or PTEX (de Koning-Ward et al., 2009), which resides at the PVM as an interface between the parasite and host environment and is supposed to play a role in unfolding and translocating the parasite proteins (Haase and de Koning-Ward, 2010; Bullen et al., 2012). The PTEX translocon is shown to consist of at least 5 components (de Koning-Ward et al., 2009), - EXP2 (a small PVM-associated protein which probably forms a membrane spanning pore, HSP101 (belongs to the ClpA/B chaperone family which probably helps in unfolding the exported proteins), PTEX150 (unknown function; Bullen et al., 2012), PTEX88 (unknown function) and TRX2 (thioredoxin-2). A recent study involving the direct growth competition assay using fluorescently-labelled *P. berghei* lines identified PTEX88 to be associated with blood stage development of the parasite as *ptex88*- mutants had a strongly reduced blood stage multiplication rates as compared to the WT parasite line (Matz et al., 2013).

In general, four distinct life forms of malaria parasite, namely the rings, the trophozoites and the schizonts, have been described to be involved in an erythrocytic developmental cycle, with a very transient fourth one, the most distinct merozoites. The parasite enters into a red blood cell in the form of merozoites, about 8-24 of which are released from a mature schizont (Lee and Fidock, 2008). A recent study has generated evidence to link the role of an epigenetic mechanism, PfSir2a (the *P. falciparum* histone deacetylase) in controlling parasite proliferation and the number of merozoites produced in a

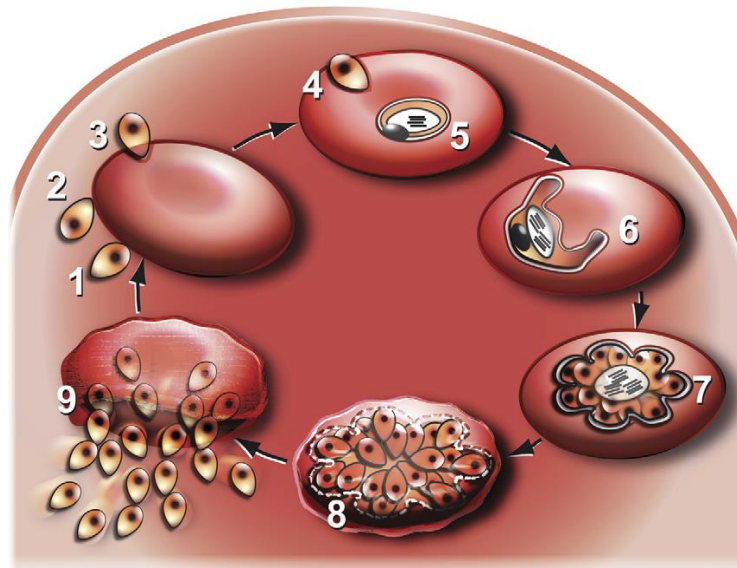
schizont. PfSir2a, which is an epigenetic silencer, was found to negatively regulate the rDNA transcription and number of merozoites per schizont (Mancio-Silva et al., 2013).

The development of the merozoites (via the ring form) into a mature trophozoite, just before nuclear division starts, takes around 16 hours. The trophozoite then matures into a schizont containing merozoites. After rupture of the mature schizonts, the free merozoites invade new red blood cells, resulting in a new erythrocytic cycle and a corresponding increase in the parasitemia (Figure 1-5).

In *P. falciparum*, one complete cycle of development of the parasite inside the RBCs takes approximately 48 hours and progresses from rings (0-24 hours) to trophozoites (24-36 hours) to schizonts (36-48 hours) finally ending in the formation and subsequent release of many merozoites from a mature schizont (Reilly et al., 2007). The total duration of the asexual blood stage development in *P. berghei* is 22-24 hours. The *in vivo* blood stage development of *P. berghei* is asynchronous as compared to a tightly synchronized *in vivo* intra-erythrocytic cycle in *P. falciparum* (Trager and Jensen, 1976). Also, *P. berghei*, *P. chabaudi* (McNally et al., 1992) and to a lesser extent, *P. falciparum* (Pasvol et al., 1980) have a preference for reticulocytes but can also invade mature red blood cells. Gross and molecular details of the invasion mechanism have now been described as a stepwise process involving initial attachment, reorientation, and invasion (Cowman et al., 2012).

The attachment of merozoites to the red blood cells is believed to be mediated by the low affinity interactions between parasite plasma membrane-anchored surface proteins and host cell receptors. The attachment is followed by a passive reorientation process mediated by a gradient of adhesive proteins (including apical membrane antigen 1 or AMA1) towards the anterior end of the merozoite (Mitchell et al., 2004, Sanders et al., 2005). A moving junction or MJ is believed to be formed before invasion which actually forms an aperture through which the parasite tries to enter the host cell (Aikawa, 1978). Through the MJ, the parasite transfers important rhoptry proteins like RON2, 4, and 5

from the rhoptry neck into the host cell to form a solid connection with the host cell cytoskeleton.



**Figure 1-5. Intra-erythrocytic development of *P. falciparum*.** 1-3: Identification, attachment, orientation and invasion of RBC by merozoite; 4: Merozoite invasion accompanied by simultaneous formation of parasitophorous vacuole membrane; 5: Ring form of the parasite characterized by large digestive vacuole; 6: Actively growing form of the parasite, the trophozoite, characterized by hemoglobin digestion and increasing accumulation of hemozoin in the digestive vacuole; 7: Schizogony, characterized by DNA replication into the syncytium; 8: Merozoites secrete axonemes in preparation to egress; 9: Rupture of membranes (PVM and RBC membrane) accompanied by release of merozoites into the extracellular space to identify fresh RBCs for initiating the next erythrocytic cycle. Source: Silvie et al., (2008) - reproduced with permission.

The established belief that actual invasion in apicomplexan parasites essentially involves an active process requiring the actin-myosin motor machinery called glideosome (Matuschewski and Schuler, 2008; Soldati-Favre, 2008; Daher and Soldati-Favre, 2009), including actin, MyoA, TRAP family proteins including MIC2 (Huynh and Carruthers, 2006) and AMA1 has recently been challenged (Meissner et al., 2013). Reverse genetics studies have shown that the invasion machinery is not essential for host cell penetration in *Toxoplasma gondii* (Andenmatten et al., 2013). These evidence leave the current understanding of the invasion process in *apicomplexa* intriguing as they suggest that there might be redundancies in the mechanism of invasion and/or

the possible existence of species-specific invasion machinery within the apicomplexans.

Whatever the machinery involved, it takes only seconds for a merozoite to invade a new red cell but this includes many steps involving multiple receptor-ligand interactions (Cowman and Crabb, 2006; Silvie et al., 2008). The merozoite has many special features related to invasion of red blood cells, such as the apical organelles, *i.e.* the rhoptries, micronemes and dense granules. Rhoptries contain proteins necessary for recognition and subsequent invasion of the RBCs. They also contain lipid stocks which are helpful in the formation of the PVM (Kats et al., 2006). The adhesins used to identify and attach the merozoite to the host RBC and also to link to actomyosin filaments for gliding and invasion are supplied by micronemes (Dowse and Soldati, 2004), whereas dense granules contain proteins for establishing communication between PVM and RBC membrane and beyond (Mercier et al., 2005).

During invasion, the contents of rhoptry bulb are secreted into the host cell thus forming the parasitophorous vacuole membrane (PVM) which enables the parasite to reside inside the host cell within its own parasitophorous vacuole (Riglar et al., 2011). Time lapse imaging has identified two clear and distinct phases of the very fast invasion process. The first or the pre-invasive phase includes the initial contact of merozoite with the RBC membrane, apical orientation of the merozoite and RBC deformation and recovery. The ensuing second or the invasion phase involves the beginning of the entry of merozoite into the red cell and its subsequent internalization. A third or post-invasive phase has also been described which includes the alterations in the shape of the RBC (echinocytosis) after the complete entry of the merozoite into the cell (Gilson and Crabb, 2009).

A number of (surface) proteins of merozoites, which are likely to play a role in invasion, appear to be conserved between rodent and human malaria parasites (Chitnis and Blackman, 2000; Cowman et al., 2000). Examples include MSP1, MSP4/5, AMA1, MAEBL, RhopH3 and SUB2.

Merozoite invasion of RBCs is central to the maintenance of parasite life cycle and malaria pathogenesis thus making it an attractive process in terms of identification of potential vaccine and drug candidates (Boyle et al., 2013). One of the recent successes in identifying such a candidate is the discovery of a single receptor-ligand pair: Basigin (BSG), an erythrocyte receptor for the parasite invasin reticulocyte binding protein 5 or Rh5 (*PfRH5*) which is believed to be required by all tested *P. falciparum* strains for entering into host red blood cells (Crosnier et al., 2011; van Ooij, 2011).

The biological complexity of *P. falciparum* is reflected in the constantly changing structure of the parasite during the repeated red blood cell cycles which are always approximately of 48 hours duration each (Bannister and Mitchell, 2003). Inside the RBC, the parasite development progresses through the following stages, based merely on the morphology of the parasite, as observed by light microscopy after nuclear staining (Arnot and Gull, 1998):

**The ring stage:** The name is derived from the typical shape of a signet ring, the parasite appears like in the Giemsa-stained blood smears after the invasion of the red cell (Bannister et al., 2000). The parasite lies in the parasitophorous vacuole (formed during the invasion process) and feeds on the hemoglobin through its cytosome. As the ring grows further, it synthesizes molecules specific to this stage and exports them to the RBC, thus modifying the red cell membrane and making it more adhesive to the non-infected RBCs (rosetting) and to the endothelium of the blood vessels (Bannister and Mitchell, 2003).

**The trophozoite:** For the next 12 hours or so, the parasite is most active in feeding and growth and becomes more rounded. New molecules are continuously synthesized and exported into the RBC - some of them assemble as membranous sacs and are seen as Maurer's clefts in stained smears. Export of new parasite proteins leads to knob formation on the RBC surface. Active feeding on hemoglobin leads to accumulation of products of hemoglobin digestion as hemozoin crystals, visible as dark pigment scattered within the food vacuole (Bannister and Mitchell, 2003). The appearance of this pigment marks the onset of the trophozoite stage (Arnot and Gull, 1998).

**The schizont:** This stage is marked by the nuclear divisions and synthesis and assembly of molecules essential for RBC invasion (Bannister and Mitchell, 2003). Schizonts are defined on the basis of the presence of more than one nucleus (Arnot and Gull, 1998). Nuclear divisions result in the formation of about 16 nuclei which move into merozoite buds around the periphery of maturing schizonts. Merozoites bud off from the residual body of cytoplasm and move into the buds. After multiple rounds of asynchronous divisions, about 8-24 mature merozoites breach the PVM and RBC membranes to escape the old red cell, for a brief extracellular phase till they eventually invade a new red cell (Bannister and Mitchell, 1998; Lee and Fidock, 2008).

**The merozoite:** Just after the release from schizont, the free merozoite is very small, ~1.2µm long and is fully armoured with the essentials of the invasion machinery - the rhoptries, the micronemes, and the dense granules (Bannister and Mitchell, 1998). This stage is immunologically important as the parasite remains in the extracellular compartment, although for a very short period of time, and thus is exposed to the host antibodies (Bannister et al., 2000). The two phases in the life of the parasite that have received an overwhelming attention of researchers are the mechanisms of release of merozoites and their subsequent invasion of the new red cell.

### 1.6.1.3 Post-erythrocytic development

**Release of merozoites / egress of parasites from RBCs:** *P. falciparum* infection is infamous for its synchronous release of merozoites from infected RBCs. The process and the molecular mechanisms involved in the infected red cell rupture and subsequent release of merozoites are still not fully unraveled (Lee and Fidock, 2008). However, recent studies reveal certain key molecular determinants involved in the parasite's egress. The whole process is understood as a highly regulated proteolytic cascade culminating in the loss of host cell integrity that preludes the release of merozoites from the RBCs. A secreted subtilisin-like protease, *PfSUB1*, is thought to play a central role in this cascade (Yeoh et al., 2007; Arastu-Kapur et al., 2008). The protease is one of the three subtilisin-like proteases expressed by *P. falciparum* and is secreted by the newly

discovered organelle - the exoneme - into the parasitophorous vacuole space shortly preceding the RBC rupture.

After its release, PfSUB1 targets and activates another family of proteases, one of the serine repeat antigens also known as SERA5 (Miller et al., 2002) and SERA6 (Ruecker et al., 2012), proteolytic activation of which is implicated in the breakdown of both, the parasitophorous vacuole membrane and the RBC membrane. In addition to its role in egress, PfSUB1 is also believed to be involved in re-invasion of fresh RBCs (Yeoh et al., 2007). A cGMP dependent protein kinase, *Pf*PKG has recently been shown to be required for the discharge of PfSUB1 into the PV and also for the release of the proteins from micronemes (Collins et al., 2013). A similar role of *Pb*PKG is established in mediating the egress of merozoites during liver stage *P. berghei* infection (Falae et al., 2010).

All these events are highly regulated and coordinated to ensure that all the merozoites contained in a schizont are fully matured prior to the release otherwise one or two of the merozoites racing ahead to maturity than others might trigger the rupturing cascade. This coordination prevents the premature rupture of PVM and RBC membrane and subsequent release of a sub-optimal number of infective merozoites (Janse and Waters, 2007). However, there are other proposed models of merozoite egress from the RBC briefly proposing that the rupture of the host RBC occurs without immediate hemolysis and that merozoites are released packaged within a thin membrane (Clavijo et al., 1998; Winograd et al., 1999; Lew, 2001).

Time lapse imaging of mature schizont stage of *P. falciparum* has revealed many details of the schizont rupture and merozoite release process (Gilson and Crabb, 2009). Just before rupture, the schizont attains full “maturity” including an increase in size, merozoites becoming increasingly distinct as they begin separating clearly and the mature schizont takes a “flower-like” appearance (Glushakova et al., 2005). As the schizont matures fully and approaches imminent rupture, the merozoites disaggregate further, presumably due to the breakdown of the PVM, and fill the RBC (Gilson and

Crabb, 2005). Thus, a fully matured pre-rupture schizont has clearly separated merozoites with no distinction of the PVM separate from the RBC membrane.

***Invasion of RBCs:*** A fresh erythrocytic cycle is initiated with the invasion of the target red cells by merozoites released from a ruptured RBC. Generally, merozoites are able to recognize new target RBCs within 1 minute of their release from their host RBCs under in vitro conditions. RBC invasion ensues and the entry into new host cell is complete on average 27.6 seconds after primary contact with the new target RBC (Gilson and Crabb, 2009).

## **1.6.2 Sexual development**

### **1.6.2.1 Gametocytogenesis**

In each asexual cycle of *P. berghei*, a small and variable proportion of parasites (5-25%) stops asexual multiplication and differentiate into sexual cells, the so-called gametocytes (Mons, 1986). It has been observed that as low as 0.2% to 1% of asexual parasites of the intra-erythrocytic phase may generate gametocytes during each cycle (Sinden, 1983). These haploid macrogametocytes (females) and microgametocytes (males) are the precursor cells of the female and male gametes, respectively. The time required for gametocytes to attain full maturity varies strikingly between rodent and human malaria parasites.

Whereas *P. berghei* (Mons et al., 1985) and *P. yoelii* (Garnham, 1966) gametocytes require 24-26 hours to reach maturity, the gametocyte of *P. falciparum* require further 8-10 days of maturation (Field and Shute, 1956) into 5 morphologically recognizable stages (I-V), following RBC invasion by a sexually committed merozoite (Hawking et al., 1971). The sex ratio in *Plasmodium* is consistently female-biased which could be explained on the basis of local mate competition theory as males in *Plasmodium* can each fertilize more than one female and an equal sex ratio would result in overabundance of male gametes (Hamilton, 1967; Reece et al., 2008). It is also suggested that sex ratios in *P. falciparum* are clone-specific and thus might have a genetic component (Burkot et al., 1984).



The genes encoding proteins secreted early from the gametocytes (up to stage II of gametocyte development), grouped into a cluster designated *Pfge* or *P. falciparum* gametocytogenesis early (Eksi et al., 2012), include Pfg27 (an RNA-binding cytoplasmic phosphoprotein and maintains cellular integrity of gametocytes; Carter et al., 1989; Olivieri et al., 2009), Pfs16 (secreted within 24-hours post-invasion; Bruce et al., 1994), Pfpeg3 or Pfmdv1 and Pfpeg4 (proteins of early gametocytes; Silvestrini et al., 2005), Pfg14.744 and Pfg14.748 (proteins exported in early gametocytes; Eksi et al., 2005; Silvestrini et al., 2010), Pfs47 (mediates immune evasion of the mosquito's immune response and is critical for transmission from a mammalian host to a mosquito; Molina-cruz et al., 2013), and Pfgeco (gametocyte erythrocyte cytosolic protein - a gametocyte exported protein; Morahan et al., 2011) among others. Among these, Pfs16, Pfpeg3/Pfpeg4 and Pfg14.744 and Pfg14.748 have been shown to be located in the parasitophorous vacuole membrane (PVM) and this commonality of origin of these proteins indicate a distinct remodelling of the PVM in gametocytes as compared to the asexual stages of the parasite (Alano, 2007). It is interesting here to note that Pfg27 has no known orthologue in other *Plasmodium spp.* other than *P. reichenowi* (a primate malaria species) thus suggesting its role in the longer developmental period of gametocytes in these species (Baker, 2010).

Sex-specific morphological features appear in the gametocytes only at 22 hours in the 26 hour of maturation of *P. berghei* (Janse and Waters, 2004) and around day 6 in the 10-day maturation of *P. falciparum* gametocytes (Alano and Billker, 2005). However, molecular differentiation of male and female gametocytes is thought to occur much earlier than the appearance of sexual dimorphism. It has been shown that Pfs16 and then Pfg27/25 are the first molecular markers for gametocytes and are expressed 24-hours into their development (Silvestrini et al., 2005). Pfnek-4 expression, in *P. falciparum*, might help in identifying the sexually committed parasites, as early as, at the schizont stage (Reininger et al., 2012) as it is known that commitment to sexual differentiation occurs at least one developmental cycle before the actual appearance of sexual-stage parasites such that all the merozoites derived from a particular “committed” schizont generate gametocytes (Bruce et al., 1990) in the following cycle and will produce all male or female gametocytes (Silvestrini, Alano and Williams, 2000).

There is a basic purpose-directed difference in mature male and female gametocytes, both at the cellular and molecular level. Males have very minimal cellular structure - reduced ribosome and endoplasmic reticulum. In comparison, the females have well developed endoplasmic reticulum, mitochondrion and apicoplast - all in preparation for the next stage of the life cycle, the rapidly developing zygote. A comparison of stage-specific proteome of *P. berghei* also identified sex-specific divergence of proteins with only 69 proteins shared between the male and female gametocytes (Khan et al., 2005). They also possess DOZI-mediated translationally repressed/silent transcripts whose rapid activation, when needed, is a key to the rapid development at the zygote stage (Paton et al., 1993; Mair et al., 2006). Osmiophilic bodies (more abundant in females) are the secretory organelles which help the gametocytes emerge from RBCs (Lal et al., 2009; Ponzi et al., 2009). Pfg377 is implicated to have a role in formation of osmiophilic bodies and gamete egress (de Koning-Ward et al., 2008). Sex-specific proteome data shows that the male proteome contains 36% of proteins unique to the sex whereas the female proteome has 19% of unique proteins (Khan et al., 2005).

Transcriptome data from several microarray experiments in *P. falciparum* suggest that about 200-300 gametocyte-specific transcripts are significantly up-regulated in the first 40-48 hours of sexual development (Eksi et al., 2005; Silvestrini et al., 2005; Young et al., 2005) including those of *P25* and *P28* which are however kept translationally repressed in gametocytes (Paton et al., 1993; Hall et al., 2005). Thus, the gametocytes appear to be arrested at the G<sub>0</sub> phase of the cell cycle in the mammalian host (Sinden, 1996). *Plasmodium* parasites are unique in the developmentally regulated expression of small subunit ribosomal RNA (SSU rRNA). There are seven of these in *P. falciparum* and four in *P. berghei*. The A-type rRNA is expressed in asexual stages whereas the S-type rRNA (previously called C-type) is expressed in gametocytes in a precursor form and processed in later mosquito stages. The switching of transcriptional activity to differential expression of the two rRNA types in *P. berghei* is associated with species-specific developmental transition from mammalian host to the mosquito. No S-type precursor forms are seen in *P. falciparum* (Waters et al., 1989; Waters et al., 1997).

There is currently no evidence suggesting the occurrence of DNA replication in male gametocytes as suggested by the absence of tritiated hypoxanthine uptake (Raabe et al., 2009) and rapid DNA replication only occurs after activation of male gametocytes in the mosquito midgut.

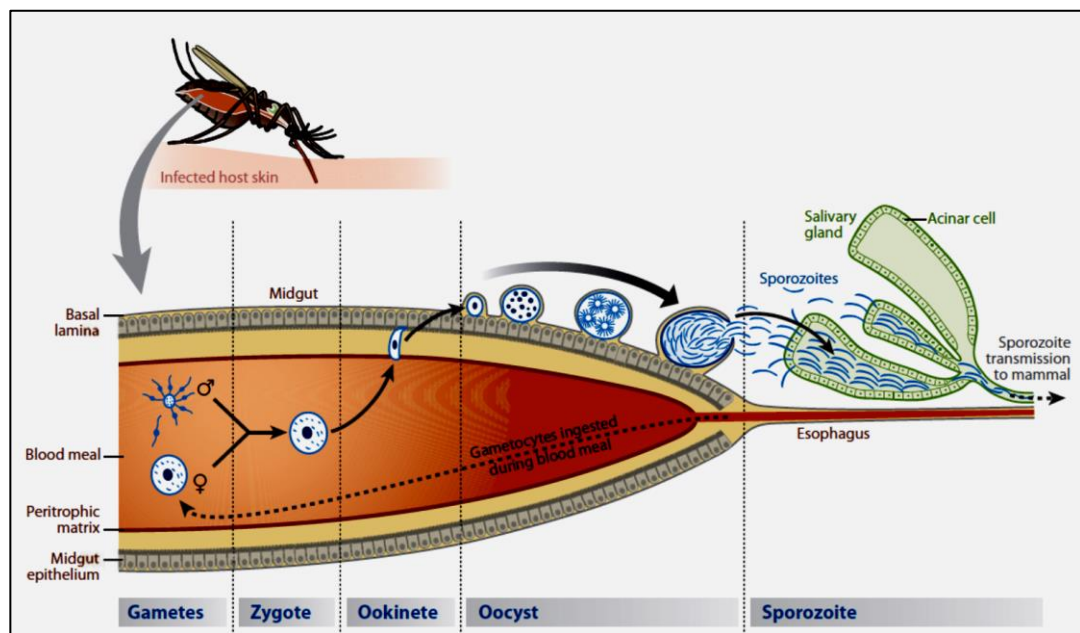
Experiments on detailing the timing of sexual commitment are difficult with *P. falciparum* as induction of gametocytogenesis by overgrowing asexual parasite cultures produces wide biological variations in microarray experiments and large-scale purification of gametocytes younger than 48 hours is difficult (Alano, 2007). Although different in many aspects from *P. falciparum* gametocytes, it is here that *P. berghei* is considered to be a desirable model to study gametocyte biology. This is further favoured by a high degree of conservation of sex-specific transcripts and proteins between the two species.

#### **1.6.2.2 Gametogenesis, Fertilization and Ookinete formation in the mosquito**

When a mosquito feeds on an infected host, only the mature gametocytes can undergo further development in the mosquito midgut. This involves an active escape of the gametocytes from the red blood cell and the formation of gametes (the process of gametogenesis). The female gametocyte differentiates into a single, spherical female gamete (macrogamete) whereas the male gametocyte produces 8 'sperm-like' microgametes which is a rapid process called exflagellation and is completed within 10-20 minutes (Figure 1-6). Male gametogenesis is associated with three rounds of DNA replication and axoneme assembly (Toyé et al., 1977; Raabe, et al., 2009) whereas female gametogenesis involves a release of the "silent" transcripts from the DOZI-mediated translational repression (Mair et al., 2006).

Three environmental triggers have been described that induce the differentiation of the gametocytes into the gametes: a drop in temperature of the infected blood to at least 5°C below that of the vertebrate host (Sinden et al., 1996), a rise in pH from 7.3 to 7.8-8.0 (Nijhout and Carter, 1978) and the presence of a gametocyte activating factors (GAF) (Billker et al., 1997; Garcia et al., 1997). This mosquito-derived GAF was shown to be Xanthurenic acid (XA) in

*P. berghei* (Billker et al., 1998) and *P. falciparum*/*P. gallenaceum* (Garcia et al., 1998). The signalling mechanisms that regulate this rapid development in the mosquito remain largely unknown. However, studies have shown a stimulation of Guanylyl cyclase (GC) by xanthiurenic acid (Muhia et al, 2001) which stimulates cGMP (Kawamoto et al., 1990) to promote exflagellation in male gametocytes in *P. berghei* and *P. falciparum*.



**Figure 1-6. Development of malaria parasite inside the mosquito vector.** The male and female gametocytes are ingested by the female *Anopheles* mosquito which activate and fertilize within the stomach to form the zygote. The zygote transforms into a motile ookinete which penetrates the midgut epithelial cells and rests under the basal lamina as oocyst. Oocyst undergoes extensive growth to generate sporozoites which when released from the oocyst travel to the mosquito salivary gland and wait for their chance of getting transmitted back to the vertebrate host. Source: Aly et al., (2009) – reproduced with permission.

It is also established that the action of cGMP in *P. falciparum* gametogenesis is mediated via a cGMP-dependent protein kinase, PfPKG (McRobert et al., 2008). A very delicate threshold-based regulation of cGMP via GC (which increases cGMP) and phosphodiesterases (PDE, which decrease cGMP) has been shown to be crucial for sexual development in *Plasmodium* (Taylor et al., 2008).

These events are also believed to activate the phosphoinositol pathway leading to a release of calcium from the cytoplasmic stores which, in turn, activates CDPK and MAP kinase, organize DNA replication, axoneme assembly, nuclear division and expulsion of gametes (Sinden, 2009; Tewari et al., 2010). A recent transcriptomic study in *P. falciparum* comparing the transcriptomes of mature non-activated gametocytes with gametocytes at 30 minutes post-activation by Suppressive Subtractive Hybridization (SSH) identified 126 genes with differential expression and 17.5% out of them were implicated to have a role in signalling mechanisms (Ngwa et al., 2013).

After approximately 10 minutes of induction to undergo gametogenesis, the gametocytes round up and emerge out of the RBCs (Billker et al., 1998). It has been demonstrated in *P. berghei* that an atypical mitogen-activated protein kinase-2 (Pbmap-2) has a role in initiating cytokinesis and axoneme motility in male gametes after the cell cycle checkpoint for completion of DNA replication and/or mitosis has been reached (Tewari et al., 2005). Between 10 minutes and 1 hour after gametogenesis, fertilization takes place in the midgut of the mosquito by penetration/fusion of the haploid male gamete to the haploid female gamete, resulting in the diploid zygote (Janse et al., 1986). It has been demonstrated that a number of gamete surface proteins are conserved between rodent and human parasites (Thompson et al., 2001; van Dijk et al., 2001). The gametes contact each other through cell-cell adhesion probably mediated by Pfs48/45 and P230. The male gamete specific P48/45, one of the 10 identified members of a protein family containing domains with 6 positionally conserved cysteine residues (the 6-cys protein family) has been shown to be responsible for recognition/attachment with the female gamete (van Dijk et al., 2001). Two other members of the 6-cys family, P47 (on female gametes) and P230 (on male gametes) have recently been identified to play a role in gamete fertilization. Pfs230 negative mutants show that the protein is essential for the characteristic agglutination of extracellular gametes with uninfected RBCs during exflagellation (Eksi et al., 2006). P47 and P230 knockout gametes have been shown to be responsible for gamete recognition and/or attachment and produce very few ookinetes with 20-50 fold reduced transmissibility to mosquitoes (van Dijk et al., 2010).

Currently, 24 proteins have been identified which have the potential to induce transmission blocking antibodies. These include P48/45, P230, P25, and P28 (Sinden et al., 2012). Pfs48/45 and Pfs230 are located in the form of a complex on the plasma membrane of the gametocyte. Upon activation, these proteins become directly exposed to the blood meal on the extracellular surface of the male and female gametes. It has been observed that monoclonal antibodies raised against these proteins blocked *Plasmodium* transmission when fed to the mosquitoes artificially (Targett et al., 1990).

Another group of six adhesive proteins called the PfCCp in *P. falciparum* (Simon et al., 2009) and lap in *P. berghei* (Trueman et al., 2004) which are expressed in gametocytes but later secreted and are supposed to play a signalling mediated role in cell-cell interactions between the female gametes during fertilization, as one of the PfCCPs, PfCCp4, has been demonstrated to associate with Pfs230 (Scholz et al., 2008).

The success of transmission of *Plasmodium* through *Anopheles* is a crucial rate-limiting step and is potentially determined by a variety of factors derived from the vertebrate host, the parasite and the mosquito (Sinden et al., 1996). One of the parasite-derived factors is a conserved male-specific sterility gene, HAP2, expressed exclusively in gametocytes and localized in the male gametocyte and the microgamete. Reverse genetics studies in *P. berghei* and complementary studies on *Chlamydomonas* have shown that HAP2 is essential not for the attachment but for the fusion of the male and female gametes (Liu et al., 2008) and thus could be considered as a potential transmission-blocking vaccine candidate (Blagborough and Sinden, 2009).

Meiotic division occurs in the zygote after about 2-3 hours of fertilization of male and female gametes (Sinden and Hartley, 1985; Janse et al., 1986). Meiosis is not directly followed by nuclear division, resulting in single nucleated zygote/ookinete with 2-4 times the haploid amount of DNA (Janse et al., 1986). The spherical zygote develops into a banana-shaped, motile ookinete within a period of 18-24 hours. The small pigment granules that are scattered throughout the cytoplasm of the gametocytes/zygotes become 'packaged' into a few clusters in the mature ookinete. It has been shown that a NIMA (never in mitosis /

Aspergillus)-related protein kinase (NIMA-related kinase or Nek-4) is critical for development of ookinete from the zygote (Reininger et al., 2005).

Ookinetes are one of the motile stages of *Plasmodium* and are known to exhibit actin-myosin-based gliding motility. Gliding locomotion and its associated machinery is shared by all members of Apicomplexa as a means of life-cycle progression as it allows them to move actively and is fast (Matuschewski and Schüller, 2008). The actin-myosin motor is typically based on scaffolds of actin and myosin polymers or myosin monomers working against F-actin tracks. The myosin involved is myosinA (MyoA) (Meissner et al., 2002), a class XIV tail-less myosin restricted to Apicomplexa and ciliates (Heintzelman and Schwartzman, 1997; Foth et al., 2006). The other components of the motility machinery includes the unique double membrane layer, the inner membrane complex (IMC) (Morrisette and Sibley, 2002), two accessory proteins, the gliding-associated proteins 45 and 50 (GAP45/GAP50) that link MyoA to the IMC through the glycoprotein GAP50 (Gaskins et al., 2004). The GAP45/ GAP50 pair likely acts in concert with MyoA-tail interacting protein (MTIP), a MyoA-associated protein that is reminiscent of the regulatory light chain (Bergman et al., 2003). Recent experiments with *P. berghei* ookinetes have strongly suggested that the Myosin A (MyoA) is vital for the motility of the ookinetes (Siden-Kiamos et al., 2011). A tight control of cGMP mediated signalling mechanisms regulated by guanylyl cyclase  $\beta$  (*PbGCB*; Hirai et al., 2006; Moon et al., 2009) and phosphodiesterase  $\delta$  (*PbPDE $\delta$* ; Moon et al., 2009) has been shown to be responsible for gliding motility and midgut invasion by the ookinetes.

### 1.6.2.3 Ookinete to Oocyst Transformation

The ookinete contains an apical complex (Sinden, 1985a) that has secretory organelles called micronemes for penetration and traversing of cells of the midgut epithelium. Several (surface) proteins of ookinetes that are involved in interaction of the ookinete with the mosquito midgut such as the peritrophic matrix (a proteoglycan matrix that separates the food from the midgut of insects), midgut epithelium and basal lamina, are conserved in structure between rodent and human parasites. Examples are Chitinase, Circumsporozoite and TRAP-Related Protein (CTRAP), P25 and P28 proteins (Yuda et al., 1999;

Dessens et al., 2001; Tomas et al., 2001). Of particular interest is the micronemal protein, MAOP (membrane attack ookinete protein) which is implicated in disruption of the host cell membrane and ookinete penetration (Kadota et al., 2004).

Other proteins that probably assist in ookinete to oocyst transformation include glycosylphosphatidylinositol (GPI)-anchored ookinete surface proteins P25 and P28 (Tomas et al., 2001), and two proteins with adhesive domains, SOAP (secreted ookinete adhesive protein) (Dessens et al., 2003) and the aforementioned CTRP (Limviroj et al., 2002; Mahairaki et al., 2005). Recently, potassium channels, particularly, have been shown to be involved in oocyst development as a knockout of such channels in *P. berghei* (*pfkch1*) was shown to have 98% reduction in mosquito infectivity as compared to the WT (Ellekvis et al., 2008).

Mature, motile ookinetes traverse the midgut epithelium by invasion of cells of the epithelium (cell traversal) and settles between the basement cell membrane and the basal lamina of the midgut wall (Han et al., 2000; Sinden and Billingsley, 2001). The invaded cells undergo apoptosis (Han et al., 2000; Vlachou et al., 2004). The ookinete invasion of the mosquito midgut is a complex process - the ookinete surface is covered by enolase which mediates the binding of plasminogen from the mammalian blood meal to the ookinete surface and this interaction, coupled with conversion of plasminogen into active plasmin, has been shown to be essential for midgut invasion and subsequent oocyst development (Ghosh et al., 2011). It is important here to note that the ookinete is the only invasive stage that is not preceded by a multiplication step, resulting in severe reduction of ookinete numbers due to activation of host protective mechanisms ignited by the invasion process (Han et al., 2000; Aly et al., 2009).

#### **1.6.2.4 Oocyst development and sporozoite differentiation**

Upon emerging from the epithelial cell, the ookinete makes contact with, but appears to be unable to penetrate the basal lamina. The cell traversal appears to trigger the change from penetrating mode to sessile mode (Aly et al., 2009). Here the parasites rapidly round up and develop into the oocyst stage.



After a growth phase of the oocyst asexual, mitotic replication results in the formation of a mature oocyst that contains thousands of daughter cells (sporozoites). A distinguishing feature of all *Plasmodium* spp. is the occurrence of a time gap between karyokinesis and cytokinesis (Aly et al., 2009). LCCL/lectin adhesive-like proteins are thought to play a key decisive role in sporozoite formation (Trueman et al., 2004).

The oocysts increase in size from 2-3  $\mu\text{m}$  in diameter to about 40  $\mu\text{m}$  (and even up to 50-60  $\mu\text{m}$ ) within 10-13 days, stretching the basal lamina overlying the oocyst (Meis et al., 1992; Aly et al., 2009). Consequently, oocysts rupture and the haploid sporozoites are released into the hemocoel that will invade the salivary glands (Sinden, 1978; Meis et al., 1989). The subsequent migration of the sporozoites to the salivary glands is poorly understood (Baton and Ranford-Cartwright, 2005). The first sporozoites reach the salivary gland at day 13-14 after the infectious blood meal. The invasion of the salivary glands by sporozoites is an active process and involves parasitophorous vacuole formation (Pimenta et al., 1994; Kappe et al., 2003). Sporozoites migrate through cells of the gland and exit into the extracellular secretory space where the sporozoites can persist for many weeks before being injected into a new host, thus completing the lifecycle (Sterling et al., 1973; Pimenta et al., 1994; Frischknecht et al., 2004).

Recent researches into the sporozoite stages of the rodent malaria parasite, *P. berghei*, have identified a family of 4 proteins, the *Plasmodium* cysteine repeat modular proteins (PCRMP1-4). The *pcrmp1* and 2 knockout parasites have been shown to be infective when injected intravenously into the host mouse but are unable to target and invade mosquito's salivary glands (Thompson et al., 2007). In contrast, the *pcrmp3* and 4 knockouts were unable to egress from the oocyst and could not enter the mosquito salivary glands. In addition, although infective to liver, the *pcrmp3-4* null mutants were unable to develop in the liver (Douradinha et al., 2011).

Several other micronemal proteins are also hypothesized to be associated with the sporozoite invasion of the mosquito salivary glands including the circumsporozoite protein (CSP), the apical membrane antigen/erythrocyte

binding-like protein (MAEBL), the thrombospondin-related anonymous protein (TRAP), and the up-regulated-in-oocysts sporozoites protein 3 (UOS3) (Mikolajczak et al., 2008), also called S6/TREP (Combe et al., 2009; Steinbuechel and Matuschewski, 2009). One of the salivary gland proteins secreted by the distal lobes of the female salivary glands and whose expression is induced by blood feeding, Saglin, is believed to act as one of the sporozoite receptors (Brennan et al., 2000; Korochkina et al., 2006; Okulate et al., 2007; Ghosh et al., 2009).

An intriguing set of proteins recently identified to have a role in *P. berghei* sporozoite development includes the *Plasmodium* Puf family of proteins which are evolutionary conserved in eukaryotes. Puf proteins typically contain a Puf domain which binds to 3'UTR of the target mRNA and leads to its translational repression and/or degradation (Wickens et al., 2002; Quenault et al., 2011). The nomenclature Puf is derived from *D. melanogaster* protein Pumilio and *C. elegans* protein fem3 binding factor (FBF) (Zamore et al., 1997; Zhang et al., 1997). Two proteins have been identified to contain the Puf domains in *Plasmodium* to date - Puf1 and Puf2. Both Puf1 and Puf2 in *P. falciparum* have been shown to be differentially expressed in gametocytes (Cui et al., 2002; Fan et al., 2004) where PfPuf2 (which is intriguingly found to be most highly expressed in sporozoites (Le Roch et al., 2003)) is believed to play a role in suppressing gametocytogenesis and differentiation of male gametocytes (Miao et al., 2010). However, in *P. berghei*, both Puf proteins have been shown to be expressed in sporozoites (Hall et al., 2005). PbPuf2 has been shown to control the sporozoite latency in mosquito salivary glands which prevents premature transformation of sporozoites before transmission to mammalian host and thus maintains infectivity to mammalian host (Muller et al., 2011). It has been recently shown that the eukaryotic translation initiation factor 2 $\alpha$  (eIF2 $\alpha$ ) kinase and a phosphatase play a regulatory role in converting salivary gland sporozoites into liver stage sporozoites (Molloy, 2010; Zhang et al., 2010).

Thus, ookinete and sporozoites are the bottleneck stages in malaria parasite life cycle and are hence critical for exploiting potential candidates for drug and/or vaccines (Aly et al., 2009). However, a paucity of *ex vivo* or *in vitro* experimental systems to study the molecular details of the events happening at

the sporozoite-salivary gland interface is a key deterrent in advancing the understanding of these mechanisms to allow the identification of drug and/or vaccine targets (Mueller et al., 2010).

## **1.7 Gametocytogenesis – a sexual developmental switch**

### **1.7.1 *Commitment to gametocytogenesis***

Since the identification and description of gametocytes which led to the discovery of *Plasmodium* parasites by Laveran in 1880 and up to the recent advances in attributing genes involved in the process of gametocytogenesis, surprisingly many aspects of gametocyte biology are still under cover (Babiker, Schneider and Reece, 2008; Dixon et al., 2008). If we take proper cues from the life cycle of the parasite, we could conveniently infer that there are certain checkpoints or bottlenecks which need to be explored further in order to have some control over interrupting the parasite life cycle. Commitment to the sexual phase of the life cycle in the form of generation of viable gametocytes (gametocytogenesis) in the vertebrate is one such essential step for establishment of infection in the mosquito vector. These processes are critical for life cycle completion and represent a major population bottleneck or checkpoint in the parasite's life (Sinden and Billingsley, 2001; Ecker et al., 2008).

Sexual differentiation in *Plasmodium* is not dependent upon segregation of sex-specific chromosomes but instead is derived from the haploid genome of the preceding asexual parasite (Cornelissen, 1988; Smith et al., 2002). In addition to the innate genetic factors, the switch from asexual to sexual stage is also probably regulated by environmental cues from the host blood (Smith et al., 2002).

Differentiation and development of sexual stages in *Plasmodium spp.* can be divided into steps of induction, commitment and maturation but the molecular basis underlying these processes are barely known (Baker, 2010). It is however known that in *P. falciparum* commitment to sexual stage occurs one

cycle before gametocytes appear in blood such that all merozoites released from a single schizont are already committed for following either a sexual cycle or an asexual one (Bruce et al., 1990; Talman et al., 2004). This implies that the trophozoites of the preceding asexual cycle were already committed to follow either a sexual or an asexual cycle (Talman et al., 2004). Also, all merozoites from a “sexually committed” schizont become either male or females (Silvestrini, Alano and Williams, 2000; Smith et al., 2000). That would indicate that commitment to male or female gametocytes either happens concomitantly or follows the asexual-to-sexual switching (Smith et al., 2002).

### **1.7.2 Factors associated with the commitment switch**

It is not known, however, when the commitment to sexual differentiation takes place in *P. berghei* parasites. The molecular mechanisms that induce and regulate the switch from asexual multiplication to sexual differentiation are still unknown. It is difficult in *P. falciparum* to study the triggers and timing of gametocytogenesis *in vivo* in humans as many inter-related factors potentially influence the switching mechanism and also because of sequestration of early gametocyte stages to spleen and bone marrow (Carter, 1988). Both in *P. falciparum* and in *P. berghei* there is evidence that the switching mechanism is highly flexible and is responsive to primary signals which include certain **environmental factors** which are believed to switch on the gene(s) that determine the commitment to the sexual phase (Dyer and Day, 2000; Alano, 2007). The role of environment as a modulator of the switching mechanism first became clear when it was demonstrated that addition of fresh blood and a lower parasitemia significantly reduced commitment of gametocytogenesis (Carter and Miller, 1979). Subsequent studies have added onto these factors which could now be grouped as those contributed by the host or the parasite, or could simply include a drug treatment or a signalling mechanism (Talman et al., 2004).

Various **host factors** contributing to increased commitment to gametocytogenesis include increased immune pressure (Smalley and Brown, 1981; Ono et al., 1986; Buckling and Read, 2001), increased steroids and corticosteroids (Lingnau et al., 1993), increased proportion of reticulocytes in blood (Trager and Gill, 1992; Trager et al., 1999) and anemia (Drakeley et al.,

1999; Price et al., 1999). Chloroquine was also found to induce gametocytogenesis (Buckling et al., 1999; Talman et al., 2004a).

Many **parasite borne factors** have been demonstrated which affect gametocytogenesis. These include the presence of mixed-genotype infections (Taylor et al., 1997; Williams, 1999), levels of asexual parasitemia (Dyer and Day, 2000) and presence of lysed parasitized erythrocytes (Schneweis et al., 1991). These environmental triggers strongly suggest a signal transduction mechanism through which the parasite receives a cue to sexual commitment. G-proteins have been implicated as a signalling mechanism to mediate the switching to sexual development in response to environmental stimulus (Dyer and Day, 2000). As there are no parasite heterotrimeric G-proteins that are homologous to mammalian G-proteins, it is possible that the parasite recruits host erythrocyte derived G-proteins for signalling (Harrison et al., 2003). The involvement of cAMP-dependent and protein kinase C-dependant signalling pathways have been shown to induce gametocytogenesis (Kaushal et al., 1980; Inselburg, 1983; Trager and Gill, 1989; Li et al., 2001; Muhia et al., 2003) however no protein kinase C have been identified in malaria parasites (Hall et al., 1997). Expression of such switch-on genes may also dictate sex specificity of gametocytes. Sexual differentiation, thus, must be dependent upon an orchestrated mechanism of gene expression and their regulation, involving sexual stage-specific gene(s) and their protein products (Lobo and Kumar, 1998).

Despite some evidence that this differentiation could also be triggered by some “secreted” factors, these have not been defined and experimentally proven until recently where it was shown that small extracellular vesicles, called exosome-like vesicles or ELVs, secreted by infected RBCs act as the mediators for transfer of DNA/signals between infected cells. These ELVs were found to be composed of both red blood cells and parasite-derived proteins. It was comprehensively shown using complementary approaches that the ELVs are internalized by other neighbouring *Plasmodium*-infected cells in the culture leading to gametocyte differentiation of the recipient asexual parasitized cells (Mantel et al., 2013; Regev-Rudzki et al., 2013). It has also been shown that ELV formation is dependent upon a protein, PfPTP2. In the context of ELV carrying sexual differentiation signals to other infected cells triggering them to commit

to gametocytogenesis, it would be intriguing to know whether deletion of PfPTP2 has any effect on the ability to generate gametocytes and what exactly is the signal carried by these ELVs (del Portillo and Chitnis, 2013; Tilley and McConville, 2013).

### **1.7.3      *Timing of commitment – when does the commitment occur?***

Despite the demonstration of fluorescent schizonts through expression of GFP labelled *SET* genes under a gametocyte-specific promoter in *P. berghei* (Pace et al., 2006), further experimentation with isolating such fluorescent schizonts and determining their gametocyte generating capabilities might prove confirmatory of the earliest molecular identification of “sexual schizonts” (Alano, 2007). An understanding of the timing of irreversible commitment to gametocytogenesis is extremely relevant to pin down the mechanisms of sexual dimorphism in *Plasmodium*.

Molecular differentiation of male and female gametocytes is thought to occur much earlier than the appearance of sexual dimorphism. It has been shown that Pfs16 and then Pfg27/25 are the first molecular markers for gametocytes and are expressed 24-hours into their development (Silvestrini et al., 2005). Pfnek-4 expression might help in identifying the sexually committed gametocytes (Reininger et al., 2012). Pfg27 knockout gametocytes were shown to be developmentally impaired with a variety of morphological forms (Olivieri et al., 2009), suggesting its role in gametocyte maturation rather than commitment. Transcriptome data from several microarray experiments suggest that about 200-300 gametocyte-specific transcripts are significantly up regulated in the first 40-48 hours of sexual development (Eksi et al., 2005; Silvestrini et al., 2005; Young et al., 2005). Experiments detailing the timing of sexual commitment are difficult with *P. falciparum* as induction of gametocytogenesis by overgrowing asexual parasite cultures produces wide biological variations in microarray experiments and large-scale purification of gametocytes younger than 48 hours is difficult (Alano, 2007).

The basic key question that is still unanswered is how the decision to switch to sexual mode from asexual cycle is induced and regulated. Further switching models describing whether the switch is constitutive and/or subjected to environmental sensing and regulation still remain debated (Sinden et al., 2012).

#### **1.7.4 Spontaneous loss of gametocytogenesis and genomic changes**

A progressive decrease in the production of sexual stages of the parasite has been observed following continuous maintenance in blood culture in *P. falciparum* (Bhasin and Trager, 1984) and during repeated mechanical passage (from one vertebrate host to another without an opportunity for vector transmission) in *P. berghei* (Mons, 1986; Dearsly et al., 1990; Janse et al., 1992b). *P. falciparum* clone 3D7 and its parental isolate NF54 are widely used gametocyte producer which during continuous *in vitro* passage lose their ability to generate gametocytes and the non-producers get positively selected (Ponnudurai et al., 1982). However, particularly in *P. falciparum*, because of the longer development of gametocytes, it happens frequently that the suggested GNP line does not reveal any gametocyte on Giemsa-staining but IFA or other high throughput techniques such as flow cytometry can detect early gametocytes suggesting that the block in gametocytogenesis occurred during the development rather than at the commitment stage (Baker, 2010).

The time taken by the parasite to lose the ability to produce gametocytes varies from a few weeks (Brockelman, 1982; Graves et al., 1984) to more than a year, both in *P. falciparum* (Carter and Miller, 1979; Trager, 1979; Ponnudurai et al., 1982) and in *P. berghei* (Janse et al., 1992b). Genetic variations have been identified within isolates of *P. falciparum* that alter the capacity to produce gametocytes (Bhasin and Trager, 1984; Burkot et al., 1984; Graves et al., 1984; Janse et al., 1989). Based on these experiments, it could be inferred that certain, possibly multiple genetic loci may control various steps in the process of gametocytogenesis and these loci become mutated and positively selected during repeated asexual multiplication and in the absence of vector transmission (Janse et al., 1992b; Lobo and Kumar, 1998). Positive selection is possible

because of the selective growth advantage offered by the non-gametocyte producer mutants. These mutations responsible for the loss in the ability to generate gametocytes might develop at several steps involved in the production of viable gametocytes (Birago et al., 1994).

The inability to produce gametocytes has been related to genomic changes in *P. berghei* including loss of repetitive DNA sequences (Birago et al., 1982; Casaglia et al., 1985) and karyotype variations (Janse et al., 1989). Large chromosomal rearrangements including a deletion on chromosome 5 in *P. berghei* (Janse et al., 1992a) and a 0.3 Mb subtelomeric deletion on chromosome 9 (Day et al., 1993; Alano et al., 1995) and a defect in chromosome 12 (Vaidya et al., 1995; Guinet et al., 1996) in *P. falciparum* have been detected and found to be correlated to the inability to produce gametocytes. The deleted region on chromosome 9 consists of 15 annotated genes - including *Pfgig* which has been characterized by gene silencing and complementation studies to be associated with reduction in gametocyte production and restoration of gametocyte-specific transcripts, respectively (Gardiner et al., 2005). Gametocyte-specific transcripts have been detected by comparing gametocyte-producing and non-producing lines (Hall et al., 2005a; Kooij and Matuschewski, 2007) and sex specific markers including *pfg377* (for females),  $\alpha$ -tubulin II (Kooij et al., 2005b) and *pfmdv-1* (for males; Furuya et al., 2005) have been identified and characterized (Silvestrini, Alano and Williams, 2000).

Recently, a forward genetics approach using PiggyBac insertional mutagenesis in *P. falciparum* identified 29 clones, called IGM (Insertional Gametocyte-deficient Mutants), which did not form mature gametocytes. Out of them, 16 were identified to be putatively responsible for gametocytogenesis. These included LCCL domain-containing (PfCCP2), Putative Hsp70-interacting protein (*PfHip*), Repressor of RNA polymerase III (MAF1), AP2 transcription factor (PF13\_0097), and Small PEXEL-containing hypothetical protein. The identification of an AP2 family protein is interesting as AP2's are already implicated in *Plasmodium* gametocytogenesis (Ikadai et al., 2013).

Although the characterization of a potential gene located on chromosome 5 which is implicated in gametocytogenesis has been attempted in *P. berghei*



(Birago et al., 1994), smaller mutations like indels and SNPs still remain undetected or unaddressed. Clones of gametocyte non-producer (GNP) lines have been developed from their parental high gametocyte producers (HP), for example, 233 (GNP) from 234 (HP) of ANKA strain of *P. berghei* (Dearsly et al., 1990) and clone HPE (GNP) from clone 8417 (HP) (Janse *et al.*, 1992b). Molecular studies involving such mutant lines defective in the ability to produce gametocytes have thrown some light in understanding the genetic control of the process of gametocytogenesis but there is still a lot to be learned and deciphered towards complete understanding of the molecular events behind these intricate processes. Comparisons of whole genomic sequences and epigenetic information between normal gametocyte producers and non-producer mutant lines could be crucial as they would help pinpoint the genetic loci controlling gametocytogenesis and identify suitable interventions.

Coupled with techniques based on molecular amplification, which are more sensitive to detect and quantify gametocytes at very low parasitemias (e.g. 0.02-10 gametocytes per microliter) (Schneider, 2006; Babiker, Schneider, and Reece, 2008), further analysis of gametocyte-specific transcripts in reference to the gametocyte non-producer lines may help refining the approximate time when the parasites have lost their ability for gametocytogenesis. However, transcript analysis is beset with the problem that the absence of any given transcript is not a reliable indicator of a causal effect, i.e. the expression of which gene or set of genes is truly responsible for the commitment to gametocytogenesis. Therefore direct whole genome sequencing methods which are unbiased representations of the differences between producer and non-producer genomes may more unambiguously lead to identification of the gene(s) responsible for commitment to sexual development.

## **1.8 Recent technological advances in Whole Genome Analysis**

### **1.8.1 *Sequencing methods***

The last 4-5 years have witnessed a major shift away from the so-called “first generation” sequencing method or the automated Sanger sequencing for genome analysis. This is partly because of the ceiling of throughput being reached with automated Sanger technology and partly due to the flurry of new or advanced sequencing techniques, known as the Next Generation Sequencing (NGS). The limitations of any further improvements in cost and throughput with Sanger sequencing generated the need to develop and improve novel sequencing technologies (Metzker, 2010). The most important achievements of the NGS are the ability to generate tremendous amount of reliable data cheaply and rapidly.

### **1.8.2 *Next-generation sequencing***

These latest and improved technologies dramatically increase the sequencing throughput by laying millions of fragmented/sheared DNA molecule on a single solid support and sequencing all these fragments in parallel, the so called Massively Parallel Sequencing (Pop and Salzberg, 2008; Tucker et al., 2009). NGS includes the commercially available technologies from Roche/454, Illumina/Solexa, Life/APG and Helicos BioSciences. Although the basic platform for NGS is similar in all these techniques, these differ with respect to the various individual processes for the analysis of a genome, such as, template preparation, sequencing and imaging, and genome alignment and assembly. Whereas 454 and Applied Biosystems platforms first attach the DNA fragments to coated beads, Solexa and Helicos attach them directly to the support (Pop and Saltzberg, 2008). Therefore, the type of data produced through each of these platforms is unique to that platform and presents specific computational challenges in the downstream processes like alignment and genome assembly. As of now, the most widely applied platform for genome sequencing is the Illumina/Solexa platform which seems to have an edge over the other currently available sequencing platforms (Metzker, 2010).

***The Illumina/Solexa platform:*** The Illumina platform is based on a clonal bridge amplification of single DNA molecules on a flow cell surface which generates 10 million single-molecule clusters per square centimetre of flow surface (Pettersson et al., 2009). Once the DNA is sheared, adapters are ligated at both ends of the fragments of DNA. These sample DNA fragments are then passed over a solid support surface which has a dense lawn of adapter complementary sequences attached to it. When the adapter-linked DNA molecules pass over this surface, each of these molecules becomes annealed to its nearby adapter-complementary sequence (primer). Bridge amplification (characteristic of this technique) is then performed which generates a double-stranded bridge after elongation (Illumina; Adessi et al., 2000). A denaturation step then frees the two strands which now remain attached to the flow cell surface by means of the adapter-complementary oligonucleotides. Repeated amplification cycles generate colony-like local clusters, each containing approximately 1000 copies and with a diameter of about 1  $\mu\text{m}$ . Sequencing is then carried out with fluorescently labeled reversible terminator nucleotides (Pettersson et al., 2009). Only one base is interrogated and incorporated at a time as further chain elongation is prevented by reversible terminators (Bentley, 2006). Once all colonies on the flow cell surface are interrogated for bases and the appropriate base attached, the fluorophores are removed and terminating bases are inactivated permitting the next cycle of base-interrogation and incorporation. The chances of mis-incorporation (errors) are claimed to be minimized by the presence of all four nucleotides during each cycle and the competition generated thereof. The relatively short read length generated by this technique may be due to incomplete incorporation of nucleotides and/or ineffective removal of reverse terminators or fluorophores (Pettersson et al., 2009).

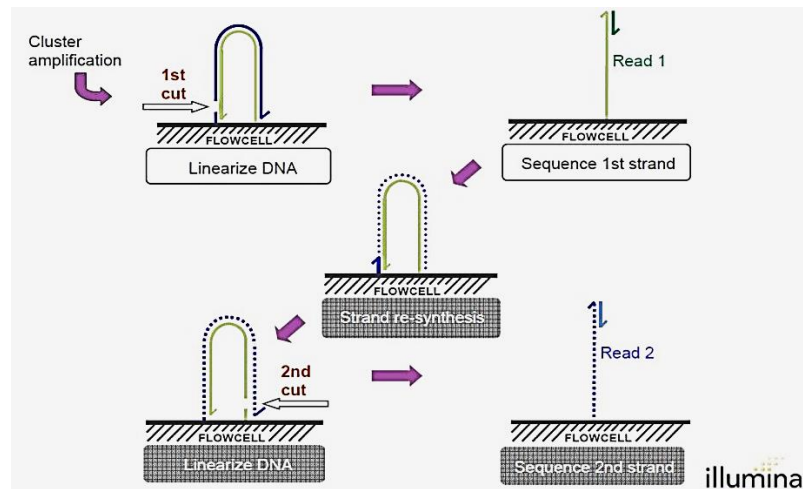
Although the length of reads generated by Illumina/Solexa platform is relatively shorter (45, 75 or 130 bases) than the average read length from Roche/454 (330 bases), the throughput is much higher (than Roche/454) in terms of base-pairs generated in each run by almost 35 and 75 fold (fragment run and mate-pair run, respectively) (Metzker, 2010). The raw accuracy is claimed to be 98.5% and the cost per base is only 1% of the cost of Sanger sequencing (Illumina website, <http://www.illumina.com>). The error rates and shorter read lengths

are compensated, at least in part, by the use of paired-end libraries (paired-end sequencing) which generate double the volume of data in each run (Pettersson et al., 2009).

All NGS platforms have this strategy for paired-end sequencing but the sequencing method from the opposite end of a DNA strand differs between various NGS platforms. Illumina uses a different hardware module and a modified paired-end enabled flow cell to re-synthesize the template using bridge amplification PCR (Figure 1-7). Sequence from one end is obtained by amplification, linearization and de-hybridization of DNA template resulting in a single strand molecule covalently attached to the flow cell which is subjected to sequencing by synthesis. A subsequent step of amplification regenerates the double-stranded template and the opposite strand is de-hybridized this time and sequenced (Holt and Jones, 2008).

### **1.8.3      *Genome assembly***

In order to create a new genome map (*de novo*) or to analyse differences between two genomic sequences (resequencing), the reads (short DNA sequences produced by NGS platforms) need to be combined and the original genome has to be reconstructed. This is accomplished by computer programs called genome assemblers using various input-specific algorithms. Major genome assembly challenges include the presence of large sections of repetitive DNA sequences (genomic repeats) distributed throughout the genome and the length of reads generated by NGS with large genomes (insufficient coverage) and/or short read lengths presenting specific computational problems (insufficiently unique sequence reads) (Mardis, 2008; Pop, 2009). The assembly of NGS data is further complicated by the bulk of the data to be assembled and moreover, by the nature of the data being generated with short(er) read lengths incorporating new(er) types of sequencing errors (Mardis, 2008).



**Figure 1-7. The technique of bridge-amplification used by Illumina-Solexa for generating the paired-end sequencing.** This method of sequencing the opposite end of DNA strand is unique to the Illumina platform. Source: Illumina website – open access.

The quality of genome assembly is found to be directly related to the read length (Chaisson et al., 2004; Whiteford et al., 2005). It has been observed that read-lengths of 200 bp or longer are necessary for good quality *de novo* genome assembly (Pop and Saltzberg, 2008). Short read lengths makes it particularly difficult to assemble specifically the repeat regions of genomes (particularly if the repetitive segments are longer than the read length) contributing to ambiguity in genome assembly and resultant fragmented assemblies (Figure 1-8). This limitation of the NGS is partly compensated by deep coverage of the genome (Krause *et al.*, 2006) and the availability of paired-end reads and mate-pair sequencing with Illumina/Solexa. Combining data generated from mate pair library sequencing with that from short-insert paired-end reads provides a powerful combination of read lengths for maximal genomic sequencing coverage across the genome and detection of structural variants (Illumina website).



**Figure 1-8. The effect of length of repeats on genomic assembly using Sanger and SRS sequencing data.** Blue: repeats longer than 800 bp. Red: repeats shorter than 800 bp. Sanger sequencing data generating read lengths greater than red repeats would result in a longer contig

represented at the top (only breaking at repeated longer than read lengths) whereas SRS (short read sequences) data would result in a much more fragmented assembly breaking at both repeat lengths. Paired-end reads (thin curved lines) tend to provide long range connectivity across repeats. Source: Pop and Salzberg, (2008) – reproduced with permission.

The length of the reads coupled with specific sequencing steps for different platforms is driving the bioinformatics community to develop new bioinformatic software and keep pace with the advancements in short read sequence (SRS) technology (Pop and Salzberg, 2008). Thus, the SRS data cannot be effectively analyzed using the software developed for the Sanger sequencing data (Nickerson et al., 1997; Moore et al., 2006; Chen et al., 2007; Huse et al., 2007) due to specific algorithmic requirements that are based on long read lengths and also to specific types of sequencing errors incorporated by SRS platforms (Pop and Salzberg, 2008).

Recently developed *de novo* genome assemblers like SSAKE (Warren *et al.*, 2007), VCAKE (Jeck et al., 2007) and SHARCGS (Dohm et al., 2007), all use the “greedy” overlap-layout-consensus logarithmic approach to tackle genome assemblies using very short sequences. Greedy algorithms join individual reads together into contigs iteratively, starting with the reads that overlap best and finishing when no more reads can be joined (Pop, 2009). The term “greedy” implies that the decisions made by the algorithm are locally optimized between any two reads and may not result in globally optimal assembly thus generating problems in assembling repeats (Pop, 2009). An alternative recent approach towards genome assembly uses the deBruijn graph paradigm (Eulerian strategy) which is less affected by the short read lengths (Chaisson and Pevzner, 2008) and thus is typically better suited for SRS data (Pop, 2009). Two recently developed assemblers incorporating the Eulerian strategy are Velvet (Zerbino and Birney, 2008) and Allpaths (Butler et al., 2008). Velvet algorithms are claimed to solve the major challenges posed by the SRS data - can remove sequencing errors and in the presence of read-pair information can resolve large repeats (Zerbino and Berney, 2008).

### 1.8.4 **Structural variants**

Variations in pathogenicity in *P. falciparum* could be attributed to the underlying genome sequence variations as detected by DNA re-sequencing methods. Such genome structural variations include single nucleotide polymorphisms (SNPs), insertion and deletion of short sequences (indels), large scale deletions, amplifications, inversions and translocations (Cheeseman et al., 2009). All variations arising out of alterations involving 2 or more base-pairs are termed as structural variations in contrast to the sequence variations which typically involve a single nucleotide (Scherer *et al.*, 2007). Gene copy number variations or CNVs (structural variants extending more than 1 kb in length) have been associated with various aspects of adaptive biology and several important phenotypes of *P. falciparum* including drug resistance (Cowman et al., 1994). Current sequence-based approaches to identify structural variants are based on paired-end read mapping (PEM) which compares the distance between paired-mate reads to the average insert size of the genomic library (Tuzun et al., 2005; Korbel et al., 2007). However, PEM based approaches have poor sensitivity to detect structural variants (SVs) in highly duplicated genomic regions and duplications larger than the insert size of the library. To rescue these limitations, newer more sensitive and accurate approaches use read depth of coverage generated by NGS methods (typically 30x) to identify CNVs (Yoon et al., 2009).

## 1.9 **Transcription and its regulation**

### 1.9.1 **General introduction**

Transcription, the cellular process of converting a particular region of dsDNA into ssRNA, is a very complex phenomenon regulated mainly by transcription factors (TFs) which either promote (as activators) or block (as repressors) the recruitment of RNA Polymerase II complex (Levine and Tjian, 2003). The transcriptional machinery has been the focus of extensive research since many years but still the finer details of how this molecular circuitry is so precisely operated and controlled are yet to be confirmed. Synthetic biology through engineering of the molecular circuitry has helped to understand how

organisms behave and how the complex eukaryotic transcriptional machinery is recruited and regulated (van Driel et al., 2003; Muller and Stelling, 2009; Khalil et al., 2012).

Gene expression is regulated at multiple levels, often called as the combinatorial regulation of gene expression. Besides regulation at transcriptional level, epigenetic and post-transcriptional regulation offer further levels of control over gene expression. Whereas epigenetics mainly involves DNA methylation and histone modifications, post-transcriptional gene regulation includes mRNA transport out of the nucleus, mRNA modifications such as 5'capping, splicing and poly-adenylation (which tends to increase the half-life of mRNA), translational repression, regulation through micro-RNAs (miRNAs), small interfering-RNAs (siRNAs) and non-coding RNAs (ncRNAs) (Chakrabarty et al., 2007; Chen and Rajewsky, 2007; Mourier et al., 2008; Zhou et al., 2010).

### **1.9.2      *Transcription Factors (TFs)***

Eukaryotic transcription factors (TFs) are intermediate protein molecules that recognize their regulatory cognate DNA sequences and then recruit the assembly of other protein complexes that control gene expression. Eukaryotic transcriptional regulation apparatus mainly includes the following: (a) RNA polymerase complex for the initiation and elongation of the transcripts, (b) the basal TFs which recruit the RNA polymerase complex and bind to the core promoter region (TSS) of the genes to be transcribed and are thus responsible for the baseline expression of the regulated gene (Ptashne and Gann, 2002), and (c) the specific TFs whose DNA binding domains (DBD) bind to the regulatory elements (cognate DNA sequence/s or the TF binding sites) in a sequence-specific fashion in the promoter region other than the core promoter to fine tune the gene expression by activating or repressing the transcription (Lodish et al., 1999).

The interactions of TFs with the polymerase complex may be direct or indirect. The indirect interaction may be mediated by chromatin remodelers or modifiers that increase access or protein-protein attractions via histone modifications (Fry and Peterson, 2001; Cosma, 2002). The interaction between



the DBD of the TF and its cognate DNA sequence is primarily dependent upon the structure of the DBD, with other protein regions of the TF trying to stabilize the complex (Kurokawa et al., 2009). Studies have shown that the DNA-binding domains of most of the TFs are highly conserved within a given phylum as compared to the rest of the protein sequence of the TF. Whereas the DBD is the principal trans-element involved in transcription initiation, the rest of the sequence might be useful in mediating protein-protein interactions and provide the activation domains (Ptashne and Gann, 1997; Setty et al., 2003; Fondon and Garner, 2004).

The sequence specificity of the interaction is largely determined by hydrogen bridges and van der Waals forces. The binding specificity of any TF can now be experimentally shown using the high-throughput Protein Binding Microarrays or PBMs which usually contain 60-mer probes in which 8-mers occur several times in different sequence context (Berger et al., 2006). Prior methods of *in vitro* detection of DNA-binding specificity of TFs included Electrophoretic Mobility Shift Assay or EMSA (Fried and Crothers, 1981; Garner and Revzin, 1981), DNase I footprinting (Galas and Schmitz, 1978), Southwestern blotting (Bowen et al., 1980) and surface plasmon resonance (Jost et al., 1991). All of these are predominantly low-throughput approaches and the level of precision for detection of a specific binding site also varies between the methods (Berger and Bulyk, 2009). The genome wide distribution of these TF binding sites can now be examined with ChIP-chip (chromatin immunoprecipitation coupled with a microarray) or ChIP-seq (chromatin immunoprecipitation coupled with next generation sequencing methods) (Johnson et al., 2007; Wold and Myres, 2008).

Sometimes, the number of identified TF binding sites far exceeds the number of expected gene targets suggesting that the binding of TFs to the TF binding sites might have roles other than direct target gene regulation (MacQuarrie et al., 2011). One possible explanation of the disproportionally and comparatively increased TF binding sites is the presence of some “non-functional” binding sites (Li et al., 2008). Presuming these are functional sites, there remains a possibility that some of these additional sites are low-affinity TF binding sites - wherein, a contact with the TF's allows baseline low level transcription of the gene sufficient enough only to allow for evolutionary

conservation purposes (Tanay, 2006). Other possibilities determining the functionality of these sites include a dependence on the concentration of the TFs, induction of chromatin looping, changes in chromatin and nuclear structure, or involvement of a novel yet undiscovered regulatory pathway/s (MacQuarrie et al., 2011).

Studies have demonstrated that the excess TF binding sites present in the intergenic regions or in repetitive elements but outside the regulatory region/s (promoters) serve as a low-affinity buffer reservoir, thus helping in fine tuning gene expression and reducing the molecular noise by limiting the concentration of the unbound/free TF (Lin and Riggs, 1975; Robertson et al., 2007; Li et al., 2008; MacArthur et al., 2009; Cao et al., 2010). In the light of the above possibilities, and many more, it would be intriguing to find out the predominant role/s played by a particular transcription factor - whether the concerned TF produces a direct transcriptional activation and/or a dominant change in chromatin landscape to regulate the transcription process.

### **1.9.3      *Epigenetic control of gene expression***

Epigenetics simply implies inheritance which is not attributable to changes in the DNA sequence or difference in external conditions (Kouzarides, 2007; Ptashne, 2007; Bonasio et al., 2010). In other words, epigenetics involves changes in gene expression without any underlying alteration in the DNA sequences (Scherf et al., 2008). The relevance of epigenetic influence through gene expression lies in the fact that comparative genome sequence analysis of mutant phenotypes alone may not reveal differences that can be causally associated to the phenotype as it may result from alterations in epigenetic marker profile or nucleosome phasing. Thus the profile of epigenetic markers and nucleosome phasing might be expected to provide useful information on genetic determinants of a particular phenotype should the genome sequence prove to be uninformative.

With the advent of so-called “sequence census” methods for measuring the whole genome “profile” including the chromatin structure, the use of hybridization assays such as microarrays for identification of protein-DNA

interactions (ChIP-chip) is not the method of choice these days (Wold and Myres, 2008). Genome-wide mapping of DNA-protein interactions are now possible with seq-based approaches such as ChIP-seq. They have an edge over ChIP-chip methods due to absence of false positive signals, more accurate quantification and a greater more dynamic resolution of signal range (Johnson et al., 2007; Wold and Myres, 2008).

Chromatin remodelling and histone tail enzymatic modifications are the two epigenetic mechanisms involved in transcriptional control. Chromatin condensation patterns determine the access of the DNA sequence to the transcriptional machinery. The densely packed chromatin or the heterochromatin is transcriptionally silent whereas the loosely packed euchromatin remains transcriptionally active.

Eukaryotic DNA is wrapped around an octamer of 4 core histones to create the nucleosome (Kornberg, 1977). Core histone proteins (H2A, H2B, H3 and H4) together with the linker histone H1 are the building blocks of chromatin. About 145-147 bp of DNA are wrapped around a histone octamer protein core to form a nucleosome chain with 10-50 bp linker DNA spacing (Teif and Rippe, 2011). Because of this tight packaging, a substantial surface area of the DNA faces the histone octamers or is spatially occluded by the neighbouring nucleosomes in chromatin fibres. In both cases, DNA is not accessible and hence not available for the TFs to bind (Luger et al., 1997; Schalch et al., 2005). The nucleosome phasing relative to the TSS has a direct effect on the transcription through the alteration of the RNA-polymerase II binding (Schones et al., 2008). Nucleosome phasing can be mapped out by using relatively straightforward digestion of chromatin with varying concentration of micrococcal nuclease (MNase) which makes double-stranded cuts on the linker DNA between nucleosome particles. This concentration dependent digestion of chromatin provides flexibility to the approach with higher concentrations of MNase producing mononucleosome-length DNA (Zaret, 2005). Variations in nucleosome spacing are likely to reflect differences in the assembly of chromatin and also the functional state of the underlying gene sequences in terms of their accessibility to the transcriptional apparatus (Blank and Becker, 1996) leading to an active transcriptional locus.

Studies have documented that all known TFs (to date) physically occupy less than a few percent of their potential cognate binding sites the remaining being not accessible. It has been further substantiated in *Drosophila* and yeast that it is the chromatin landscape which basically dictates where an individual TF must bind by controlling the accessibility of the other potential target sites (Carr and Biggin 1999; Iyer et al., 2001; Yang et al., 2006; Joseph et al., 2010; Kaplan et al. 2011). The TFs must find a way to expose these hidden DNA fronts for transcription to start (Zaret and Carroll, 2011).

Chromatin looping is another chromatin-mediated phenomenon which basically provides transcriptional control by bringing the distant regulatory elements into functional proximity of the target genes (Ragoczy et al., 2006; Schoenfelder et al., 2010; Yochum et al., 2010). Studies in yeast have shown that TFs binding to the genome-wide recognition sites might not help in regulating the gene expression directly but may alter the chromatin and/or nuclear structure so that the accessibility of these regions is sufficiently enhanced to allow the direct interaction of downstream TFs expressed later in development (Badis et al., 2008; Hartley and Madhani, 2009; Ganapathi et al., 2011).

The mechanism underlying this epigenetic control through histone tail enzymatic modification includes covalent post-translational modifications on over 30 residues of histone tails generally of histones H3 and H4 (Strahl and Allis, 2000; Grant, 2001). These modifications predominantly include acetylation, phosphorylation, and methylation but also deimination, ubiquitylation and sumoylation, ADP ribosylation, histone tail clipping, etc. (Bannister and Kouzarides, 2011). In general, whereas histone acetylations (mediated by histone acetyltransferases or HATs) have been found to be associated with transcriptional activation (by relaxing the chromatin thus promoting transcription) and DNA repair (Masumoto et al., 2005); methylations (mediated by histone methyltransferases or HMTs) control the formation of transcriptionally active and inactive regions of chromatin and usually establish a silent chromatin domain blocking transcription (Grant, 2001; Hakimi and Deitsch, 2007).

### 1.9.4 ***The promoter region of the gene***

The promoter is the region upstream of the coding sequence wherein the TF-RNA polymerase complex binds and initiates transcription. The TFs are assembled in the pre-initiation complex (PIC) by the transcription adapter molecules. The PIC then places the RNA-polymerase II complex at the transcription start site (TSS) in the basal or core promoter located normally within 50 bp upstream to the TSS (Kornberg, 2007). The rest of the promoter might offer TF binding sites (TFBS) which provide fine tuning of the transcription process. This is typical of classical bi-partite eukaryotic promoters comprising a basal promoter region and an upstream regulatory region. In eukaryotes, the presence of TFs is critical for transcription to start and hence the transcription is switched off in eukaryotes by default (Wray et al., 2003). The TFs are believed to control transcription through a combination of actions at the control regions including promoters, enhancers, silencers and insulators, which might be located on either side of the TSS. Enhancers and silencers respectively promote and prevent transcription directly whereas insulators tend to prevent the action of enhancers on the promoter.

The effect of the distance between the binding sites of repressor and activator proteins on gene expression was highlighted recently (Fakhouri et al., 2010) wherein it was shown that the repressor efficiency was higher at smaller distances between the repressor and activator binding sites present on the gene promoter region. Novel gene expression within an organism may arise through a combination of changes in upstream regulators (the *trans* factors) and mutations in noncoding regulatory DNA sequences of a gene (the *cis* factors) (Rabeiz et al., 2011). Further, the long stretches of “junk” genomic DNA might contain recognition sequence for certain TFs (Berman *et al.*, 2004). Any random point mutations and/or indels in these regions might also produce a functional regulatory region (Stone and Wray, 2001), further complicating the regulation of gene expression in eukaryotes.

## 1.9.5 **Regulation of gene expression in *Plasmodium***

### 1.9.5.1 **Introduction**

*Plasmodium*'s complex life cycle between two hosts involves well defined morphological stages accompanied by a masterly orchestrated stage-specific gene expression profile, the mechanisms regulating which are broadly unknown (Coleman and Duraisingh, 2008). The parasite's gene expression follows a typical eukaryotic expression pattern but with an exclusive stage-specific well-coordinated expression profile in a highly, almost 90%, AT-rich genome (Gardner et al., 2002; Balaji et al., 2005; Hall and Carlton, 2005). The whole 48-hour parasite transcriptome looks sigmoid, signifying successive expression of genes, also called the "just in time" expression profile (Bozdech et al., 2003a; Llinas et al., 2006). The following may be factors contributing to this, including, but not limited to - high levels of anti-sense transcription (transcription complementary to other RNA transcripts), unique patterns of mRNA storage and decay, stage-specific and well-orchestrated gene expression cascades, relative paucity of established TFs and regulatory elements and poorly defined gene promoter regions (Aravind et al., 2003). Genome analyses however revealed a relative paucity of both *cis*- and *trans*-regulatory elements for transcription in *Plasmodium* which reflected/suggested a minor role of regulation in transcription by the TFs (Coleman and Duraisingh, 2008; Llinas et al., 2008).

### 1.9.5.2 **Transcription factors in *Plasmodium***

Evolutionarily speaking, the expansion of TFs across the eukaryotes shows a lineage specific pattern wherein the core TFs are mostly conserved throughout the eukaryotes but the distribution of specific TFs across eukaryotes is more or less unique for each lineage (Chervitz et al., 1998; Riechmann et al., 2000; Lespinet et al., 2002). Despite this lineage-specific expansion of the specific TFs, the *apicomplexa* showed an intriguing apparent deficiency of known, conserved TFs as evident by the significantly (10-25 fold) higher ratio of the total number of genes to the total number of TFs as compared to free-living yeasts (Templeton et al., 2004).

This could be explained partially by the parasitic life of the *apicomplexa* but then the explanation is overruled by the facts that the phylum possesses an extensive network of structural and regulatory chromosomal and signalling proteins (Aravind et al., 2003; Templeton et al., 2004), has a very complex alternating two-host life-cycle and demonstrates an intricate but well-orchestrated stage-specific gene expression profile (Bozdech et al., 2003; Le Roch et al., 2003). From this paradox, it was inferred that either the phylum *apicomplexa* had an undetected machinery of specific TFs which are unrelated or distantly related to the known families of DNA-binding domains and/or possessed an alternate regulatory mechanism/s, other than the TFs, including epigenetic and post-translational regulation (Balaji et al., 2005).

The debate was answered (in part) by the discovery of a specific clan of lineage specific TFs in *apicomplexa* through sensitive sequence profile analyses methods. This new family of TFs possessed one or more copies of the DNA-binding domains related to plants' Ethylene Response Factors (ERF) or Apetala-2 domain or more popularly, the AP2 domain. The striking stage-specific differential expression profile of ApiAP2 genes suggested their potential role in regulating transcription in apicomplexans (Balaji et al., 2005). Even with the discovery of ApiAP2 as specific TFs in *Plasmodium*, the ratio of genes to specific TF for *Plasmodium* would not be any closer to that for eukaryotes which would imply that either there might be more specific TFs yet undiscovered in *Plasmodium* or the regulation of gene expression is even more complex and involves a variety of post-transcriptional regulators.

Certain other DBD-containing proteins have been documented to be involved in regulation of key genes related to cell cycle regulation and progression; however their definitive DNA binding has not yet been shown by Chromatin Immunoprecipitation (ChIP). These include **PfMyb1** which is known to bind to the promoters of phosphoglycerate kinase, calcium-dependent kinase, TATA-binding protein, proliferating cell nuclear antigen, histones, and cyclin-dependent kinase (Doerig et al., 1995; Doerig et al., 1996; Boschet et al., 2004; Gissot et al., 2005) and **high-mobility-group (HMG) box proteins** which are effective inducers of pro-inflammatory cytokines like TNF $\alpha$  and inducible Nitric oxide synthase (iNOS) and thus are thought to be useful in producing host

inflammatory immune response in malaria (Briquet et al., 2006; Kumar et al., 2008). Studies involving *P. falciparum* in culture and *P. berghei* in mice suggest that the *Plasmodium*-infected red blood cells stimulate the host mononuclear cells to release HMGB which induces the secretion of host TNF $\alpha$  (Higgins et al., 2013). HMG box proteins have also been shown to be involved in controlling genes important for oocyst development in mosquito (Gissot et al., 2008).

Certain TFs have also been shown to bind to intergenic regions. This could be partly explained by the extreme (>80%) AT-richness of the *Plasmodium* genome (Gardner et al., 2002) which renders certain physical properties to the DNA such as increased curvature, more stable hydrogen bonding patterns (Nelson et al., 1987) and less efficient packaging into nucleosomes (Balaji et al., 2005; Polson and Blackman, 2005). In short, apart from Myb and HMG box proteins, the recent discovery of ApiAP2 family of TFs as evolutionary conserved proteins that bind to at least one cognate DNA motif (Eckert et al., 2005; Painter et al., 2011), have provided a deeper insight into the regulation of stage-specific gene expression in *Plasmodium* (Yuda et al., 2009; Yuda et al., 2010; Iwanaga et al., 2012).

### **1.9.5.3 Epigenetic regulation of gene expression in *Plasmodium***

Gene expression and coordination in *Plasmodium* spp. as in so many eukaryotes is believed to involve multiple layers of regulation including epigenetic control (Hakimi and Deitsch, 2007; Dzikowski and Deitsch, 2009). The existence of the phenomenon of epigenetic control of gene expression is indirectly supported by the apparent lack of families of recognizable transcription factors in *Plasmodium* spp. (Hakimi and Deitsch, 2007; Scherf et al., 2008). These parasites possess the full machinery of chromatin-remodelling proteins which strongly suggests that epigenetics has a role to play in gene expression (Miao et al., 2006; Navadgi et al., 2006; Cui et al., 2007; Cui et al., 2008; Iyer, et al., 2008). Research has shown that epigenetics in *Plasmodium* is majorly involved in regulation of transcription, delimitation of functional genomic elements and antigenic variation (Cortes et al., 2012).



In *Plasmodium*, tri-methylation of histone 3 lysine 9 (H3K9me3) has been shown to be responsible for gene silencing and di- or tri-methylation of histone 3 lysine 4 (H3K4me2/me3) and acetylation of histone 3 lysine 9 (H3K9ac) responsible for activating transcription (Scherf et al., 2008; Fischer et al., 2008; Lopez-Rubio et al., 2009). DNA sequence elements that encode epigenetic modifiers such as histone acetyltransferases (HATs), deacetylases (HDACs) and methyltransferases (HMTs) have also been identified in *Plasmodium spp.* (Aravind et al., 2003; Horrocks et al., 2009). This epigenetic marking is dynamic and changes throughout the parasite asexual cycle (Salcedo-Amaya et al., 2009).

Direct assessment of chromatin activity at or near *var* genes TSS has shown enrichment of typical epigenetic marks (Lopez-Rubio et al., 2007; Fischer et al., 2008; Scherf et al., 2008; Salcedo-Amaya et al., 2009; Cui and Miao, 2010) associated with gene silencing (histone 3 lysine 9 tri-methylation or H3K9me3) (Duraisingh et al., 2005; Flueck et al., 2009; Perez-Toledo et al., 2009) and activation (di- or tri-methylation of histone 3 lysine 4 or H3K4me2/me3 and acetylation of histone 3 lysine 9 or H3K9ac) (Freitas-Junior et al., 2005). In this regard, the regulation of *var* gene expression has been studied in detail and the major *var* gene regulon was found to be composed of a *var* gene promoter and its intron (Deitsch et al., 2001; Duraisingh et al., 2005; Freitas-Junior et al., 2005; Voss et al., 2006; Perez-Toledo et al., 2009), the ApiAP2 TF PfSIP2 (Flueck et al., 2010), the histone deacetylase duo PfSIR2A and PfSIR2B (Tonkin et al., 2009), the H3K4 methyltransferase PfSET10 (Volz et al., 2012), heterochromatin protein 1 (PfHP1) and the histone variant H2A.Z (Petter et al., 2011).

A very recent study with *P. falciparum* has convincingly established the role of one of the SET-domain bearing histone lysine methyltransferases, PfSET2 in regulating the H3K36me3. PfSET2 was shown to be responsible for keeping the remaining 59 odd *var* genes silenced (hence it is also called PfSETvs; for *var* gene silencing) barring the one which is activated at any given time. Each of these *var* genes encode an individual PfEMP1 which has a role in *P. falciparum* virulence and it is this *var* gene switching that is the possible immune evasion mechanism adopted by the parasite inside the host. When PfSETvs was deleted, all the *var* genes were expressed together corroborating its role in *var* gene silencing.

Further, RNA- Fluorescent in-situ hybridization (FISH) experiments confirmed that all the var transcripts studied could be colocalized at a particular locus of the nuclear periphery - a specific transcriptionally active site (also shown by Lopez-Rubio et al., 2009). The study also showed that H3K36me3 occupancy at a silent var gene was considerably higher compared to the active one (Jiang et al., 2013). Table 1-1 shows a composite summary of various histone modifications and proposed functions as observed in *Plasmodium*.

#### **1.9.5.4 Nuclear structure and gene regulation**

FISH studies in *P. falciparum* have revealed a specific perinuclear arrangement of 60 var genes (of subtelomeric and more centromerically located clusters) in the nucleus instead of random distribution and the presence of a particular sub-nuclear var gene expression site. Although precise mechanisms of gene expression control in relation to the spatial occupation of active transcription sites within nucleus are not known, further studies have shown that there exists multiple active transcription sites both at the nuclear periphery (Ralph et al., 2005; Voss et al., 2006; Lopez-Rubio et al., 2009) and at the centre (Freitas-Junior et al., 2000; Moraes et al., 2013). A recent study showed the difference in spatial organization of transcription sites in *P. falciparum* rings & trophozoite stages - the transcription foci were found to be located in the outermost nuclear region in rings as compared to an even distribution in trophozoites (Moraes et al., 2013).

Table 1-1. Histone modifications in *Plasmodium*

Histone	Position	Amino acid	Modification	Notation	Enzyme	Proposed Functions	
H2A	N-terminal	--	Acetylation	H2A-N-term-ac			
	3	Lysine	Acetylation	H2AK3ac			
	5		Acetylation	H2AK5ac			
H2B	112	Lysine	Ubiquitination	H2BK112ub			
H3	4	Lysine	Methylation	H3K4me		Permissive euchromatin (di-& tri-Me), transcription activation	
			Methylation-di	H3K4me2			
			Methylation-tri	H3K4me3	PfSET1		
			Methylation-tetra	H3K4me4			
	9		Acetylation	H3K9ac	PfSET3	Transcription activation	
			Methylation	H3K9me			
			Methylation-tri	H3K9me3	PfGCN5	Transcription silencing (tri)	
	14		Acetylation	H3K14ac			
			Methylation	H3K14me			
			Arginine	Acetylation	H3R14ac	PfGCN5	
				Methylation	H3R14me	PfGCN5	
	17		Lysine	Methylation	H3K17me	PfCARM1	
		Arginine	Methylation	H3R17me			
			Methylation-di	H3R17me2			
	18	Lysine	Acetylation	H3K18ac			
	20		Methylation-tri	H3K20me3			
	23		Acetylation	H3K23ac			
	27		Acetylation	H3K27ac			
	36		Methylation-di	H3K36me2	PfSET2	Transcription silencing (tri)	
			Methylation-tri	H3K36me3	PfSET2		
	56		Acetylation	H3K56ac			
	79		Methylation-tri	H3K79me3			
H4	N-terminal		--	Acetylation	H4-N-term-ac		
	3		Arginine	Methylation	H4R3me	PfPRMT1	
		Methylation-di		H4R3me2			
	5	Lysine	Acetylation	H4K5ac	PfMYST		
			Methylation	H4K5me	PfMYST		
			Acetylation	H4K8ac	PfMYST		
			Acetylation	H4K12ac	PfMYST		
			Methylation	H4K12me			
			Acetylation	H4K16ac	PfMYST		
	17	Arginine	Methylation	H4R17me			
	20	Lysine	Methylation	H4K20me	PfSET8	Transcriptional silencing, heterochrom-atin (tri)	
			Methylation-di	H4K20me2			
			Methylation-tri	H4K20me3	PfSET8		

**Table 1-1. Summary of some of the known histone modifications in *Plasmodium*.** Source: The table is modified, adapted and reconstructed to suit the needs from Trelle et al., 2009; Cui and Miao, 2010 and from the company Epigentek® and is used with permission.

### 1.9.5.5 Post-transcriptional regulation of gene expression

The regulation of gene expression after transcription is believed to play a major role in controlling transcription in *Plasmodium* (Le Roch et al., 2004; Shock et al., 2007; Foth et al., 2008; Agarwal et al., 2011). Based upon the transcriptomic studies in *Plasmodium*, it was established that almost 80% of the total genome remains transcriptionally active during various life stages signifying successive expression of genes, again following the “just in time” expression profile (Bozdech et al., 2003; Llinas et al., 2006). This supports the hypothesis that the protein expression is also controlled post-transcriptionally (Bozdech et al., 2003), that is, specific regions in the transcripts themselves regulate the stage-specific translation of the mRNA (Arbeitman et al., 2002), for example, U-rich regions in the UTRs act as translational repressors in *P. berghei* and are important for zygote development (Braks et al., 2008). Certain transcripts in female gametocytes have been shown to be kept in the repressed state mediated by a DEAD-box RNA helicase or DOZI until the translation of the repressed transcripts is needed (Mair et al., 2006; Braks et al., 2008; Mair et al., 2010).

Other examples of post-transcriptional gene regulation includes mRNA modification through 5'capping (Shuman, 2002), splicing (Zhang et al., 2011) and poly-adenylation (Shock et al., 2007; Narayan et al., 2009). The mRNAs are known to acquire a 7-methyl guanosine (m7G) cap in the nucleus which is important in nuclear export and translational initiation (Shuman, 2001; Gu and Lima, 2005; Sonenberg and Hinnebusch, 2009). Detection of uncapped mRNA species in *Plasmodium* strongly suggests that it might be a strategy in the parasite for mRNAs to remain translationally repressed until needed (Mair et al., 2006; Shaw et al., 2007). Polyadenylation at the 3'UTR of the mRNA also exerts a regulatory role in gene expression (Oguariri et al., 2006; Shock et al., 2007).

Post-transcriptional regulation of gene expression in *Plasmodium* is also mediated by splicing - the process of joining two exons by removal of intervening intron/s through small nuclear RNAs or snRNAs (Francoeur et al., 1985; Upadhyay et al., 2005) and alternative splicing by generating premature stop codons in the spliced variants (Otto et al., 2010; Sorber et al., 2011).

## 1.9.6 *ApiAP2 Transcription factor family in Plasmodium*

### 1.9.6.1 Introduction

ApiAP2 family of TFs are a relatively newly discovered family of trans-acting proteins in *Plasmodium* with a proven role in stage specific gene expression during the ookinete, sporozoite and liver stages of the parasite. Since there is no known eukaryotic lineage-specific expansion of the AP2 family other than that in *apicomplexa*, this specific lineage of AP2 proteins in *apicomplexa* is termed as *apicomplexa* AP2 or the ApiAP2 family (Balaji *et al.*, 2005; Iyer *et al.*, 2008; Painter and Llinas, 2012). Currently, the ApiAP2 family is the only lineage-specific TF family identified in *Plasmodium spp.* (Iyer *et al.*, 2008).

### 1.9.6.2 Conservation and functional significance

ApiAP2 TFs have been found throughout *apicomplexa* including *Theileria*, *Cryptosporidium* and *Toxoplasma* (Balaji *et al.*, 2005; Altschul *et al.*, 2010; Iyer *et al.*, 2010). Although well conserved across the *Plasmodium spp.*, two of the ApiAP2 proteins tend to be species-specific; PFL1075w in primate malarias and PVX\_080355 in *P. vivax* and *P. knowlsei*. This species-specific conservation might indicate important specific role played by these TFs in controlling functions restricted to the involved species such as hypnozoite formation (Painter *et al.*, 2011). And because of the lack of similarity of the ApiAP2 family of proteins to the human proteins, these are priority candidates for drugs and vaccines (van Ooij, 2008).

Microarray experiments and RNA-seq analyses have found that the majority of the 27 ApiAP2 proteins were found to be temporally transcribed during the asexual intraerythrocytic development cycle (IDC) of the parasite (Bozdech *et al.*, 2003; Le Roch *et al.*, 2003; Otto *et al.*, 2010) and some are transcribed during the sexual stages of the parasite thus it can be inferred that ApiAP2 TFs are important regulators of gene expression throughout the *Plasmodium* life cycle stages (Silvestrini *et al.*, 2005; Young *et al.*, 2005; Tarun *et al.*, 2008; Yuda *et al.*, 2009).

### 1.9.6.3 Structure of ApiAP2 proteins

The family contains 27 AP2 DBD containing proteins in *P. falciparum* and these are highly conserved throughout *apicomplexa* (Bischoff and Vaquero, 2010). The AP2 domains in apicomplexans are found both as single domain as well as tandem domain arrangement i.e., two tandem AP2 domains separated by a short linker amino acid sequence (Balaji et al., 2005). However, some members of the ApiAP2 family contain three DNA binding domains (PFF0670w, PF10\_0075, PF11\_0404 and PF13\_0235) (Balaji et al., 2005).

All the members are known to be expressed throughout the intraerythrocytic developmental cycle of the parasite (Balaji et al., 2005; Painter et al., 2011). Protein binding microarrays (PBM) and EMSA's have established that the DBDs recognize a specific stretch of DNA sequence/s (target motif/s) in the promoter region of the gene whose transcription is to be controlled and bind to it/them in a sequence-specific manner. The motifs for AP2 domain in *Plasmodium* are more AT-rich as compared to the classical plant AP2 domains. The binding domain architecture and its sequence-specificity for binding to the target motif are also shown to be highly conserved between orthologous pairs of AP2 domains of even phylogenetically distant apicomplexans such as *Plasmodium falciparum* and *Cryptosporidium parvum* despite sequence variations in the rest of the pairs of ApiAP2 proteins. This specificity is possibly mediated through the conserved predicted  $\beta$ -strand residues. Although the DNA-binding specificity is conserved, the putative target genes are not conserved between distant *apicomplexa* (DeSilva et al., 2008).

Apart from the presence of DNA-binding domains, the rest of the protein family is unremarkable in terms of the presence of other domains and/or motifs for targeting the protein to specific sub-cellular localizations except for the presence of classical lysine- and arginine-rich nuclear localization signals supporting their role as TFs (DeSilva et al., 2008). Studies have also suggested that AP2 proteins interact with each other and also with other molecules including the chromatin remodelling factor, *Plasmodium* histone acetyltransferase, GCN5 (PF08\_0034), at least *in vitro* (LaCount et al., 2005), a high mobility group (HMG) protein, fork head domain protein, and a plant

homeodomain (PHD) containing protein in *Plasmodium* (Bougdour et al., 2010). A part of PfSIP2 (PFF0200c; Meissner et al., 2005) and ApiAP2-O (PBANKA\_090590; Pino et al., 2012) has been shown to act as a transactivation domain in yeast system and has been assayed in tetracycline transactivator-based inducible conditional knockout vectors for *P. berghei* (Krizec and Sulli, 2006).

#### 1.9.6.4 Structural determinants of ApiAP2-DNA binding

The crystal structure of one of the DNA-bound dimers of PfAP2s DBD, PF14\_0633, shows that the  $\beta$ -sheet fold binds the major groove of the DNA via base-specific contacts with support from the  $\alpha$ -helix. Based on the requirement of two copies of the recognition motif for high affinity binding of the AP2-O with its cognate DNA, it was hypothesised and later shown that the ApiAP2 proteins hetero-dimerize when bound to their target DNA (Lindner et al., 2010). Dimerization may be important as it has been shown to influence gene regulation (Amoutzias et al., 2008). No additional Pfam domains have been predicted for any of the AP2 proteins although they have been hypothesised to contain activation domains and/or domains involved in protein-protein interactions which have not yet been identified (Finn et al., 2008; Lindner et al., 2010).

The domain dimerizes via a 3D domain-swapping mechanism in which the  $\alpha$ -helix of one protein is packed against the  $\beta$ -sheet of its dimer counterpart. This dimerization allows the Cys76 residues of each monomer to interact with one another permitting disulphide bond formation between the two (Lindner et al., 2010). This binding induced dimerization may also bring two distant DNA loci close together (a conformational rearrangement) in the nucleus and/or may induce functional rearrangements conducive to transcriptional regulation. It is also established that during the base-specific contact of the domain with the cognate DNA, certain bases such as Asn72, Arg74, Arg88, and Ser90, which make base-specific contacts with the motif, are conserved throughout Apicomplexa and hence appear critical for the contact (Lindner et al., 2010).

#### 1.9.6.5 DNA-binding specificities

Exhaustive biochemical and computational characterization of the entire *P. falciparum* AP2 family proteins revealed that multiple DBDs within the same protein bind to distinct DNA motifs for many AP2 proteins. Even proteins with a single DBD have secondary recognition motifs in addition to the primary ones which might be associated with reducing expression noise and fine-tuning of gene expression together with a widening of the choices of target genes that could be regulated by a particular TF (Campbell et al., 2010).

#### 1.9.6.6 ApiAP2 target genes

Bioinformatic detection of 5' upstream promoter regions of *Plasmodium* genes enriched for the specific DBD motif have been used to identify potential downstream target genes for certain ApiAP2. For PFF0200c (PfSIP2), apart from the individual ApiAP2 specific target genes, intriguingly, the target motif array (consisting of a direct (T/G)GTGC(A/G) repeat spaced by four nucleotides) was found to be enriched in the promoter regions of *var* gene families (Flueck et al., 2010), other ApiAP2 genes and even on the same ApiAP2 gene indicating the many different ways in which these TFs regulate the expression of other genes including other TFs and also themselves i.e., auto-regulation (De Silva et al., 2008). PfSIP2 is highly conserved although the *var* gene family is species-specific and not present in all *Plasmodium* species - the reasons for this are currently not known.

However, it should be noted that the mere presence of a recognition motif does not guarantee an active interaction which has to be proven through competent techniques for detecting protein-DNA interactions including, but not limited to, chromatin immunoprecipitation or ChIP (Collas, 2010; Cai and Huang, 2012; Furey, 2012).

#### 1.9.6.7 Role of ApiAP2 family in *Plasmodium* gene regulation

The AP2 family of TFs have been shown to be the only class of sequence-specific transcription factors with an established role during various life cycle



stages of *Plasmodium spp.* (Balaji et al., 2005; Iwanaga et al., 2012). What is common about the three AP2 TFs characterised recently (see below) is that these are involved in regulating the stage-specific gene transcription at various parasite progression points, for example, AP2-O (PBANKA\_090590) at the ookinete stage (Yuda et al., 2009), AP2-Sp (PBANKA\_132980) at the mosquito sporozoite stage (Yuda et al., 2010) and most recently, AP2-L (PBANKA\_021440) at the liver sporozoite stage (Iwanaga et al., 2012). It is important to note here that these TFs were identified in *P. berghei* as the rodent malaria system allows the characterization of these genes at the respective life stages providing a significant advantage as compared to the *P. falciparum* system. Also, because these genes are not essential to blood stages which cannot be studied in all models, they could be characterised.

#### **1.9.6.7.1     *ApiAP2-O***

A systemic examination of the AP2 genes not expressed in the asexual parasite life stages of the rodent malaria parasite, *P. berghei*, revealed the AP2 protein called AP2-O (PBANKA\_090590; orthologue of PF11\_0442) which later proved to be essential for the formation of the invasive ookinete stage of the parasite life cycle, thus providing the first direct evidence of the role of AP2 TFs in regulating stage-specific gene expression in *Plasmodium* (Yuda et al., 2009). The *AP2-O* has orthologs in other *Plasmodium spp.* as well and encodes a protein containing 1258 amino acids which has a single highly conserved AP2 domain. It has been shown that AP2-O is highly expressed in the ookinete stage and is subjected to DOZI (Development of Zygote Inhibited) mediated translational repression (Mair et al., 2006; Yuda et al., 2009). *AP2-O(-)* parasites showed an inability to infect mosquitoes and when grown in culture showed an abnormal morphology of the ookinetes (pear-shaped as compared to banana-shaped in WT). *AP2-O(-)* ookinetes also demonstrated a 5-fold decrease in expression of 15 genes, of which 4 were ookinete stage-specific genes. EMSA's clearly documented that the AP2-O DBD recognizes and preferentially binds to the 6 bp sequence (TAGCTA) in the promoter region of the genes in addition to binding with similar sequences, although with a weaker affinity. *In vivo* experiments confirmed that TAGCTA sequence worked as a *cis*-acting element which was conserved across the *Plasmodium spp.* (Yuda et al., 2009; Campbell et al.,

2010). AP2-O is also shown to be expressed during the intraerythrocytic developmental cycle thus suggesting that it might have additional (non-essential) roles to play other than in ookinete development (Bozdech et al., 2003; Le Roch et al., 2003; Otto et al., 2010).

#### **1.9.6.7.2      *ApiAP2-Sp***

A similar ESTs search for the AP2's in the *P. berghei* mosquito sporozoite-stage AP2 database in PlasmoDB resulted in identification of PBANKA\_132980 or AP2-Sp (PF14\_0633). The expression of *AP2-Sp* starts from the late oocyst stage and continues up to the salivary gland sporozoite stage in the mosquito. *AP2-Sp* (-) parasites grew well up to the oocyst stage in the mosquito but did not generate the invagination of the plasma membrane necessary for the formation of the sporoblasts. Electromobility shift assays or EMSA's clearly documented that the AP2-Sp DBD recognizes and binds to the 8 bp sequence (TGCATGCA) in the promoter region of the genes in a sequence-specific manner. Further refinement of the recognition motif identified that the minimum sequence essential for binding is TGCATG. Furthermore, when the AP2 DBD of endogenous AP2-O was swapped with that of AP2-Sp, the ookinete stage demonstrated induction of AP2-Sp target genes (Yuda et al., 2010). This would mean that the specificity of downstream gene activation depends purely upon the specific DBD/recognition sequence pair and once activated the subsequent effects are mediated by general/shared co-factors that are ubiquitously expressed (at least in sporozoites and ookinetes). AP2-Sp is documented to be expressed throughout the intraerythrocytic developmental cycle of *P. falciparum* and at least during the trophozoite stage of *P. berghei* (Bozdech et al., 2003; Le Roch et al., 2003; Hall et al., 2005).

#### **1.9.6.7.3      *ApiAP2-L***

A survey of expressed sequence tags (ESTs) coding for the AP2's in the *P. berghei* liver-stage AP2 database in PlasmoDB resulted in identification of multiple ESTs for PBANKA\_021440 or AP2-L. The *AP2-L* has orthologs in other *Plasmodium spp.* and encodes a protein containing 1272 amino acids which has 2 highly conserved AP2 domains. *AP2-L*(-) parasites showed 10,000-fold reduced

liver infectivity as compared to the WT. *AP2-L(-)* sporozoites were cultured in HepG2 cells and stained with anti-CSP antibody to show that the nuclear division of the *AP2-L(-)* liver stage parasites was arrested at 36 hours post-infection. This mid-schizont arrest caused the observed decrease in the liver cell infectivity of the *AP2-L(-)* parasites (Iwanaga et al., 2012).

#### 1.9.6.7.4 *PfSIP2*

Recently, it was discovered that one of the ApiAP2 members, PfSIP2 (SPE2-interacting protein; PFF0200c), binds to the SPE2 DNA motif array (consisting of a direct (T/G)GTGC(A/G) repeat spaced by four nucleotides) present in the sub-telomeric regions located upstream to certain *var* genes (Voss et al., 2003; Flueck et al., 2010). Other than the DNA mediated actions of these TFs, it has also been shown that the ApiAP2 proteins do interact with each other and also with the chromatin machinery, including the *Plasmodium* histone acetyltransferase GCN5 (LaCount et al., 2005). When PfSIP2 was overexpressed, it had no effects on the global transcriptional profile suggesting that it might not be involved in transcriptional regulation directly but instead has a role in formation and maintenance of heterochromatin (Flueck et al., 2010). The reasons why PfSIP2 is highly conserved across *Plasmodium spp.* which do not have *var* genes are not known currently.

#### 1.9.6.8 Combinatorial gene regulation and ApiAP2 TFs

Evidence has now been made available regarding the role of ApiAP2 TFs in regulating gene expression in combination with epigenetic machinery. Chemical perturbation of class I and II histone deacetylase inhibitors in blood stage *P. falciparum* parasites by apicidin was found to be associated with increased expression of some of the ApiAP2 genes (which would normally be down-regulated otherwise) (Chaal et al., 2010). This is supported by evidence of enrichment of H3K9me3 marks at the *pfl1085w* ApiAP2 gene locus which is associated with silencing of *pfl1085w* (Lopez-Rubio et al, 2009; Salcedo-Amaya et al., 2009). These findings suggest a considerable overlap in the mechanisms involved in regulation of gene expression in *Plasmodium spp.* including the ApiAP2 TFs, epigenetic control and post-transcriptional control through

translational repression. These combinatorial and overlapping gene regulation mechanisms in *Plasmodium* are completely understandable in the light of its complex yet well-orchestrated stage-specific gene expression profile.

#### 1.9.6.9 Summary and future perspectives

All the current research evidence on ApiAP2 TFs have established them as the prime regulators of stage-specific gene expression at all developmental stages in *Plasmodium* parasites (Balaji et al., 2005) with some evidence gathering from other Apicomplexans as well including *Theileria* spp. (Marta Pieszko and Brian Shiels, personal communication). Further studies are definitely needed to detail the molecular mechanism by which so few TFs (predominantly ApiAP2 family) act to tightly orchestrate complex gene expression pattern of *Plasmodium* spp. (Tuteja et al., 2011). Another intriguing area of research is the need for multiple DBDs within a TF together with the functional significance of unremarkable (other than presence of nuclear localisation signals) regions outside the DBD's which might underpin the complexity and/or redundancy in transcriptional regulation by this family of TFs (De Silva et al., 2008).

Further, research is also needed to exploit these TFs as candidates for drugs as they carry huge potential towards becoming the next line anti-malarial agents of the future, keeping in view the ongoing spread of resistance against the current front line anti-malarials, the ACT. Also, because AP2s affect multiple downstream genes it could be more difficult for the parasites to develop resistance against them (Tuteja et al., 2011). If ApiAP2s can be developed and used as drug targets, they will surely be among the best candidates to achieve the daunting goal of global malaria eradication in the coming decades, unless some new transmission blocking drug candidate is discovered.

## 1.10 *Plasmodium berghei* – the rodent malaria parasite

### 1.10.1 *Model malaria parasites*

Model malaria parasites provide unique insights into the developmental biology of *Plasmodium* and its interactions with host and vector. The African (thicket) rat parasites including *Plasmodium berghei*, are popular models for the study of human malaria (Carlton et al., 2005). The combination of *in vitro* cultivation of *P. berghei* (particularly cultures of liver and mosquito stages) and *in vivo* infections permit the detailed analysis of the phenotype of the parasites including gamete fertilization, zygote development, synchronous blood stage development, and gametocyte formation (Leiden Malaria Research Group website - <http://www.lumc.nl/con/>). Also, their usefulness in reverse genetic studies is well established (de Koning-Ward et al., 2000) as the most straightforward and efficient approach in any *Plasmodium* species (Janse et al., 2011).

An inherent and recognized limitation of malaria observational studies in humans is that studies of association cannot prove causality with the result that human studies on malaria are sometimes considered inconclusive. Even with such recognized limitations, human studies on malaria are needed but lack of access to human organ/tissue samples make it necessary to include animal malaria models (Langhorne et al., 2011). Thus, rodent malaria models, particularly, *P. berghei* have been extremely valuable for understanding parasite biology and studying host-parasite interactions in malaria. Considering the limitations of each approach, human malaria studies and rodent model malaria studies, it has been found useful that the two approaches should be attempted not in competition to each other but rather in complementation so that they each inform the other (Langhorne et al., 2011).

The availability and suitability of a wide range of efficient reverse genetic technologies for *P. berghei* with the possibility of analysing parasites throughout their complete life cycle, both *in vitro* and *in vivo*, have made *P. berghei* the

most frequently used malaria models for gene function analysis (Carvalho and Menard, 2005; Janse et al., 2006). The ease of completing the entire life cycle of the parasite under laboratory conditions, including both the vertebrate stages and the mosquito stages, further offers a great opportunity to study the complex parasite biology in many aspects (Balu and Adams, 2007; Kooij et al., 2012).

With almost 80% of the 5300 or so genes in the *Plasmodium* genomes being orthologous, the relevance and significance of studies involving *P. berghei* becomes compelling (Janse et al., 2011). Comparative genomic studies of *P. falciparum* and *P. berghei* have revealed a core set of ~4,500 *Plasmodium* orthologues located in the highly syntenic regions of chromosomes (Kooij et al., 2005a) which implies a high degree of conservation of genome organization, content and possibly gene regulation between *P. falciparum* and *P. berghei* (Janse et al., 1994; Rich and Ayala, 2003).

Although favourite research models for human malaria, rodent malaria models have been criticized for some of their limitations. These include bypassing the liver stage of the parasite life cycle during infection of the mouse which is usually achieved by intravenous or intraperitoneal routes. However, the liver stage infections are achievable through the infected mosquito feeding on anesthetized mice which has now become a routine procedure for establishing evidence for studies involving sexual stages of the rodent malaria parasites.

Also, the lab mouse is now gradually being supplemented with the natural host, the African thick-knee rat. Secondly, as the laboratory mouse is not the natural host of the rodent malaria in the wild, intricate differences in the host-pathogen interactions may not be ideal or representative of the wild type scenario. This is particularly relevant in immunological studies related to vaccine development as deciphering the protective immune response to blood-stage *P. falciparum* infections in humans has been difficult (Doolan et al., 2009). There is then the third way to link evidences between rodent malaria and human malaria experiments by generating transgenic rodent malaria parasites containing an orthologue of the endogenous gene of interest from the human malaria parasites (Sinden et al., 2012).

## 1.10.2 Genetic engineering techniques in *P. berghei*

### 1.10.2.1 Transfections in *P. berghei*

Stable transfection and gene targeting with a drug selectable marker was reported first in *P. berghei* (van Dijk et al., 1995; van Dijk et al., 1996) followed by *P. falciparum* (Wu et al., 1996) and much later in another rodent malaria parasite, *P. yoelii* (Mota et al., 2001). *P. berghei* is considered to be the best malaria model for successful stable transfection due to the long and tedious procedure to transfect *P. falciparum* combined with a low transfection efficiency of the order of  $\sim 10^{-6}$  (O'Donnell et al., 2002).

The exogenous DNA used for transfection has to cross at least three different membranes in *P. falciparum* in order to be able to reach the parasite nucleus (the RBC membrane, the parasitophorous vacuole membrane and parasite cell membrane) which makes transfections less efficient in *P. falciparum*. *P. berghei*, on the other hand offers the advantage of using merozoites for transfection thus providing easier access for the exogenous DNA.

A distinct advantage in *P. berghei* is the ability to transfect a linearized DNA construct which drastically reduces the chances of development and maintenance of circular plasmids as episomes as observed in obligatory transfections with circular DNA constructs in *P. falciparum* (Kadekoppala et al., 2001; O'Donnell et al., 2001; Thathy and Menard, 2002). The negative selectable marker available for use in *P. falciparum* (HSV/tk; herpes simplex virus thymidine kinase) is also not very specific in selection, thus killing non-hsv/tk expressing parasites as well (Duraisingh et al., 2002). The ease of completing the entire life-cycle of *P. berghei* under lab conditions is a big advantage in terms of studying the complex biology of the parasite.

High efficiency transfection: The use of non-viral nucleofector® transfection technology (Amaxa GmbH) increases the transfection efficiency to the order of  $10^{-2}$ - $10^{-3}$  (Sakamoto et al., 2005; Janse et al., 2006). Such high transfection efficiency would mean lesser time to select the transfectants and easier selection (sorting) of fluorescently-labelled transfectants using FACS even in the absence of drug selection. Recent developments in *P. berghei* transfection

experiments like the small-scale *in vitro* culture protocol (Somsak et al., 2011) using as little as 20 µl of mouse tail blood has been shown to be equally efficient for a single transfection and at the same time offering multiple advantages over the conventional large-scale culture and purification protocol in terms of reduced requirement of time, parasite sample, consumables and most importantly, animals.

Positive-Negative drug selection: The use of a single positive-negative drug selectable marker cassette in the *P. berghei* transfection construct has significantly improved the flexibility in reverse genetics approaches. Use of a positive selection marker, such as hDHFR (offering resistance to the drug pyrimethamine) and a negative selection marker such as the fusion protein from yeast cytosine deaminase and uridyl phosphoribosyl transferase fusion protein (yFCU; offering sensitivity to the drug 5-FC) technically allows the possibility of achieving phenotypic rescue in single crossover mediated gene disruption (Braks et al., 2006), and selectable marker recycling or vector recycling for subsequent genetic manipulation of the same line. An improved negative selection method which replaces the need of i.p. injection of 5-FC with providing 5-FC dissolved in drinking water reduces the number of animals required for the procedure and other indirect costs while maintaining the consistency and reliability of the method (Orr et al., 2012).

### **1.10.2.2 Studying essential genes in *P. berghei***

With the inability to use gene knockdown or silencing approaches, such as RNAi, in *Plasmodium* due to the absence of the necessary machinery (Baum et al., 2009), the following developments have opened a new area for studying the functions of genes which are essential for the parasite and hence cannot be knocked out.

- (a) Conditional knockout system: A conditional mutagenesis system using the Flp/FRT site-specific recombinases from yeast (Branda and Dymecki, 2004) has been developed in *P. berghei* (Carvalho et al., 2004) and successfully applied (Lacroix et al., 2011). A modified cre/loxP based conditional mutagenesis system, called the DiCre system has recently



been successfully used in *P. falciparum* for site-specific recombinase mediated gene deletion. The DiCre system has a great potential for use in *P. berghei* which will further expand the *P. berghei* reverse genetics toolkit (Collins et al., 2013).

- (b) Transport shuttle mutagenesis: using a Tn5 transposon derivative in *Escherichia coli* (Sakamoto et al., 2005), it provides an indirect way to identify essential genes in blood stages by screening for genes that cannot be disrupted as the genes already disrupted using transposons are directly transformed into *P. berghei*.

### 1.10.2.3 Gene complementation in *P. berghei*

The process of genetic complementation (functional restoration of null mutant to wild type phenotype by introducing a fully functional WT copy of the disrupted gene) opens opportunities to establish a causal association between an implicated gene and its hypothesized function. Particularly, as in Apicomplexa, where an estimated ~47% of the gene pool comprises of hypothetical genes and/or genes with unknown functions (Florent et al., 2010), complementation experiments may serve as powerful tools in understanding the molecular basis of a variety of basic and applied phenomena in Apicomplexan biology (Striepen et al., 2001; Sultan et al., 2001; Florent et al., 2010). Despite the use of appropriate controls, it may be difficult to refute that the resulting phenotype from gene targeting experiments is not due to accumulation of independent genomic alterations during the recombination events or due to the presence of the altered recombinant / genomic locus *per se* (Janse et al., 2011). Looking at the vital role of complementation in establishing the causal inference, it is highly recommended that complementation experiments should become a part of standard procedure to causally associate genes with their functions (Goldberg et al., 2011).

However, in *Plasmodium* species, complementation of gene knockouts is not widely attempted due in part to the inherent difficulty in cloning of large coding regions of such AT-rich genomes in order to be able to complement the full length of the knocked out gene, particularly when the gene to be

complemented is large. Also, low transfection efficiency in *P. falciparum* is another deterrent for successful complementation in the human malaria parasite. Although highly recommended, complementation can be substituted by the isolation of two independent mutant clones exhibiting the same phenotype in view of the above difficulties in complementing the gene. The role of a second independent mutant is said to be indispensable in difficult to complement genes (Goldberg et al., 2011).

#### **1.10.2.3.1 Complementation techniques**

Basically, all complementation techniques are aimed at introducing the wild type copy of the disrupted gene back into the *Plasmodium* genome. Depending upon the method used to disrupt the gene (single/double crossover homologous recombination, or a natural/induced mutation(s) involving SNP(s)/indel(s)) and the extent of gene disruption (complete or partial), the DNA complementation construct and thus the method used for selecting the complemented parasites would vary.

Because the single/double crossover-based gene disruption will ideally have the drug selectable marker cassette (SMC) already integrated into the genome at the endogenous gene locus, it is mandatory to use the positive-negative SMC which is not recyclable (that is, must not contain two copies of the homologous 3'UTR in the SMC which could potentially mediate vector recycling either spontaneously or under negative selection drug pressure) for gene disruption. A double crossover-based complementation of the WT copy of the gene at the endogenous locus (using appropriate length of the gene based on the extent of gene disruption) could then be attempted using negative selection to select the complemented parasites from the transfectants.

In cases where the disruption was mediated by SNPs or indels, the endogenous locus will not have a SMC integrated (until otherwise shown) and the complementation could be attempted with a construct bearing the full length WT copy of the gene flanked by sufficiently large homology arms to enable double crossover-based integration of the construct into the genome. One of the homology arms, preferably the downstream or the 3' homology arm, should bear the SCM in order to be able to select transfectants with the complemented copy

of the gene using positive selection. The 5' homology arm should be spared for genetic manipulation as the presence of any extra (exogenous) DNA here might interrupt the promoter of the gene of interest thus interfering with the functional restoration of the gene to be complemented. This is particularly important in *Plasmodium* wherein the precise length of the promoters is largely unknown.

It is important here to keep in mind that any attempt to complement the gene would be extremely difficult if the mutant to be restored has acquired a phenotypic growth advantage (as in the case of gametocyte non-producer mutants generated in the current study) as a result of the disruption. Even if strong negative selection based approach is used, there remains a biased likelihood that the natural selection pressure of the mutant parasites will overcome the drug selection pressure thus jeopardizing the complementation efforts.

Recent application of complementation has shown the successful use of negative selection based approach (the Gene-In Marker-Out or the GIMO approach) to introduce the WT copy of the gene (complete or partial) thus replacing the disrupted endogenous locus. The complementation construct can simply be a PCR product of appropriate length amplified using a proof reading polymerase. The GIMO approach claims to offer multiple advantages over conventional complementation approaches in terms of being faster, less laborious and requires fewer animals (Lin and Annoura et al., 2011). Additionally, GIMO method could be clubbed with the recently developed “recombineering” based high-throughput gene targeting approach (Pfander et al., 2011).

Apart from introducing the WT copy at the disrupted endogenous locus, complementation could also be achieved by introducing and expressing the WT copy of the gene (transgene) at another dispensable locus in the genome, such as the p230 paralog locus in *P. berghei* (Janse et al., 2006). Ideally the gene at the exogenous locus should be expressed under its own endogenous promoter and the 3'UTR.

#### 1.10.2.4 Recent advances in *P. berghei* genetic engineering techniques

Some of the recent advances in *P. berghei* genetic engineering techniques include:

pBAT plasmids: Recently, highly optimized and versatile gene targeting vectors were developed for *P. berghei* genetic manipulations (Kooij et al., 2012). Collectively called as pBAT (*Berghei* Adaptable Transfection) plasmids, these are claimed to be suitable for vector recycling, robust GFP labelling of recombinant parasites, C-terminal mCherry epitope tagging, and heterologous gene expression. With small size and versatile restriction endonuclease recognition sites, these plasmids promise a high throughput approach to *Plasmodium* reverse genetics.

PCR-generated knockout constructs: A two-step PCR-based method for generating the linearized transfection construct for double crossover recombination in *P. berghei* has been now devised (Ecker et al., 2006). The technique offers significant improvements over the conventional *E. coli* based generation of constructs in terms of bypassing the difficulties of cloning AT-rich DNA into *E. coli*, time saved over the lengthy and cumbersome conventional method and improving the yield of the construct. The PCR-based approach could have limitations in generating larger constructs due to the inherent difficulties in error-proof PCR-based amplification of larger DNA.

PiggyBac insertional mutagenesis: Forward genetic approaches have not been widely applied in *Plasmodium* research because of the lack of adequate tools for whole genome analysis. However, recently insertional mutagenesis using PiggyBac transposable elements (widely used functional genomics tool in many species) was applied to *P. berghei*. The piggyBac system has been successfully adapted and applied for *P. falciparum* forward genetics (Balu et al., 2009). Such transposon-mediated random mutagenesis which widens the genetic manipulation armamentarium in *P. berghei* can be used as forward genetic screens for analysing gene function, particularly, for the many 'hypothetical' *Plasmodium* proteins (Fonager et al., 2011).

### 1.10.2.5 Recombineering based constructs for genetic manipulation of *P. berghei*

Larger than conventional (0.4-1.0 kb) homology arms (Janse et al., 2006) may be required for homologous recombination-based targeted genetic interventions at certain times to improve the recombination efficiency and to minimize the chances of random integration into the genome, particularly when the gene of interest is sufficiently large and when the complementation is to be attempted against a selection pressure favoring the disrupted genotype. Long homology arms, of up to 10 kb, have been shown to increase the recombination frequency substantially in other model systems (Hasty et al., 1991). However, cloning of such larger pieces of highly AT-rich DNA sequences (Gardner et al., 2002; Hall et al., 2005) has been shown to be difficult in the traditional *E. coli* system as larger AT-rich inserts become unstable in *E. coli* and hard to manipulate using restriction-ligation based molecular cloning. To circumvent such difficulties, the recombineering strategy (Pfander et al., 2011) has been developed for *P. berghei* to generate gene targeting constructs bearing larger than conventional homology arms thus avoiding the creation of large circular plasmids for the purpose.

A *P. berghei* gDNA library for large scale genetic manipulations is not available, for reasons stated above, in bacterial artificial chromosomes (BAC) and hence such a library was developed in low copy number plasmid based on bacteriophage N15, which replicates in *E. coli* as a linear, double stranded DNA molecule (Ravin and Ravin, 1999; Ravin et al., 2003) which is able to stably carry AT-rich *Plasmodium* sequences (Godiska et al., 2010). Such large AT-rich genomic DNA inserts in the N15 bacteriophage are amenable to homologous recombination in *E. coli* that transiently express the recombinase complex and proofreading activity encoded by the bacteriophage lambda red $\gamma\beta\alpha$  operon and bacterial *recA* under the control of the arabinose-inducible pBAD promoter (Zhang et al., 1998; Wang et al., 2006). This approach, called lambda red recombineering or simply recombineering is robust and independent of restriction sites and is thus the method of choice to scale up targeted gene disruption in the mouse (Skarnes et al., 2011).

The recombineering strategy has been successfully applied to convert the gDNA library inserts into gene deletion, tagging and complementation vectors for the genome-wide functional analysis of *P. berghei* genes. Basically, a bacterial drug SMC was first introduced into the gDNA insert, either deleting the target gene thus making the construct useful for targeted gene knockout; or placing the SMC within either of the UTRs of the gene making the construct suitable for complementation studies; or preparing for 3'-end tagging. To complement a gene disrupted locus already bearing the SMC (see above), the large genomic inserts in the N15 bacteriophage could directly be used after linearization with appropriate unique restriction endonuclease. The bacterial marker was later replaced with a *P. berghei* positive-negative selection cassette in a Gateway LR Clonase reaction *in vitro*. NotI unique restriction sites placed at either end of the modified library insert could be used to release the construct from the plasmid backbone making it ready for transfection (Pfander et al., 2011).

### **1.10.3     *The genome of P. berghei:***

Comparative genomic information and sequencing of additional *Plasmodium* genomes will provide in the understanding of the inter- and intra-species diversity in malaria parasites, the biology of the parasite and its interactions with mammalian host and mosquito vector (Pain and Hertz-Fowler, 2009). It is evident that the recent technological advances in DNA sequencing together with falling costs of the techniques have rendered relatively facile the complete sequencing of the small genomes that are typical of apicomplexan parasites (~25Mb) (Carlton et al., 2005; Pain et al., 2005). Similarly, advances in computing power and algorithm design mean that such genomes can be assembled accurately on desk top computers. The same technologies can also be harnessed to perform global epigenetic analyses, examining the genome-wide distribution of histone modifications. The small size of a *Plasmodium* genome means that such characterisation can be achieved on a dynamic basis and the changing patterns of histone modification can be assessed and compared across distinct developmental stages.

### 1.10.4 Challenges in assembly of *P. berghei* genome

*P. berghei* has an estimated genome size of 25-27Mb with 14 chromosomes (Kooij et al., 2006; Wellcome Trust Sanger Institute website - [http://www.sanger.ac.uk/Projects/P\\_berghei/](http://www.sanger.ac.uk/Projects/P_berghei/)). The assembly of rodent malaria parasite genomes has lagged behind that of the human parasites but the genome of *P. chabaudi* (strain AS) is now 97% complete and that of the closely-related *P. berghei* (clone ANKA and line 233 believed to be derived from ANKA) is advanced towards completion. As of February 2010, the *P. berghei* genome comprises 4881 predicted proteins (5124 in *P. chabaudi*), 22 pseudogenes (3 in *P. chabaudi*), 121 gaps with 68 gaps in core regions (3 gaps in *P. chabaudi*), and 43 unassigned contigs (23 in *P. chabaudi*). Fifty eight genes have been annotated in the unassigned contigs, all of them belong to subtelomeric gene families. The genome is expected to be completed shortly (Chris Newbold, Sanger Institute, Personal communication).

The sub-telomeric regions are presenting particular problems in completing the *P. berghei* genome assembly. In the *Plasmodium* genome, the sub-telomeric regions (extending 10-40 kb towards the internal part of the chromosomes) consist of species-specific repeats (Carlton et al., 2005). Extensive chromosome size polymorphism, exclusively involving differences in copy number of a largely universal 2.3 kb sub-telomeric repeat unit, has been observed during *in vivo* multiplication in *P. berghei* (Pace et al., 1987; Dore et al., 1990; Ponzi et al., 1990; Janse and Mons, 1992a). The complex and variable number of repetitive DNA in the sub-telomeric regions of the chromosomes still presents a remarkable challenge towards *P. berghei* genome assembly closure. The sub-telomeric regions are known to be particularly unstable (Pace et al., 1995) and they also share extreme similarity between chromosomes (Carlton et al., 2002; Gardner et al., 2002) which makes it difficult to map the scaffolds of contigs to the respective chromosomes.

## 1.11 Aims and objectives

This research was started with the aim of exploring the hypothesis that the loss of ability to produce gametocytes is reflected in the organisation of the parasite genome at the level of the genome sequence and/or organisation as reflected by changes in the genome-wide epigenetic profile and/or nucleosome phasing. The hypothesis was investigated using deep sequencing technology however, despite the indication and necessity of performing the epigenetic and nucleosome profiling; they could not be employed due to shortage of time.

Broad aims of the thesis:

1. To assess the global changes in a *Plasmodium* genome that are associated with the loss of the ability to commit to sexual development (produce gametocytes).
2. To provide a causal link between the observed genomic changes and gametocytogenesis.



## **2 Materials and Methods**

## 2.1 General Parasitological methods

The rodent malaria parasite *P. berghei* completes its natural life cycle in the wild in two hosts (Chatterjee et al., 2001) - asexual cycle in mammalian host, the thicket rat (*Grammomys surdaster*) and sexual cycle in mosquitoes (*Anopheles durenii*). Under laboratory conditions, the complete life cycle of the parasite was maintained either completely *in vivo* between Theiler's Original (TO) mice and mosquito (*Anopheles stephensi*) or in a combination of *in vivo* and *in vitro* conditions. Completion of the full *P. berghei* life cycle is currently not possible *in vitro* probably because of the very low rates of spontaneous release of the merozoites from mature schizonts and subsequent re-invasion *in vitro* (Jambou et al., 2011).

### 2.1.1 Parasite growth *in vivo*

#### 2.1.1.1 Animal handling and maintenance

All animal experimentations were performed in strict accordance with the Home Office regulations and UK Animals (Scientific Procedures) Act 1986. The animal experimentation protocols were approved by the University of Glasgow Ethics Committee. A personal license (PIL 60/12380) was obtained from the Home Office to carry out regulated procedures on mice. Adult female outbred Theiler's Original (TO) mice weighing 20-30g and aged approximately 6 weeks were procured from Harlan Laboratories (Shardlow, England, UK) and maintained in animal house for experiments.

#### 2.1.1.2 Infecting the mice with parasites

Desired inoculum of the parasites was introduced to the mice either through:

- (a) Intra-peritoneal (i.p.) injections on one or either side of the flanks of unanesthetized or Isoflurane (3-5% in Oxygen) anesthetized mice. This route of infection was used when infecting the mouse with blood stage parasites from cryopreserved stabilates or when mechanically passaging the parasites from an infected mouse to a naïve mouse. Briefly,

a cryotube containing ~0.5 ml of blood in equal volume of Glycerol/PBS solution (30% Glycerol v/v containing 50 µl Heparin stock (200 IU/ml in RPMI1640) solution) was taken out of the liquid Nitrogen storage tank and thawed at room temperature. A total of 0.1 to 0.5 ml of the suspension was then injected i.p. into the flanks of the mice with a sterile hypodermic 1 ml syringe. For mechanical passage, a drop of tail blood (~5 µl) from an infected mouse (parasitemia between 5% and 15%) was collected into 1 ml of PBS in a microcentrifuge tube and 0.1 to 0.5 ml of the diluted blood was then injected i.p. into the flanks of the mouse,

(b) Intra-venous (i.v.) injections in the tail vein of unanesthetized or Isoflurane (3-5% in Oxygen) anesthetized mice. Intravenous route was used for infecting the mice with the blood stage parasites for cloning the parasites, or synchronizing the infection or for transfection (see details below). Prior to the i.v. injections, the mice were put into a warm box at 30 °C for a period of ~30 minutes so that the tail veins were dilated. Once sufficiently dilated, approximately 0.2 ml of the infected blood suspension was injected into the tail vein with a sterile hypodermic 0.5 ml insulin syringe. During the i.v. injections, the mice were either kept steady with the help of a restrainer or isoflurane-anesthetized (3-5% in Oxygen),

(c) The bites of infected female anopheline mosquitoes (*A. stephensi*) under anaesthesia. This route was used to assess the infecting capacity or transmissibility of the parasites from infected female anopheline mosquitoes to the uninfected mouse. For such transmission bite-back experiments (TBB), naive mice were injected with Hypnorm® (Fentanyl-fluanisone; Crown Chemicals) / Hypnovel® (Midazolam; Roche Diagnostics) anesthetic combination approximately 30 minutes prior to the mosquito feed. The Hypnorm®/Hypnovel® mix was prepared by adding 1 part (0.15 ml) each of the two drugs to 2 parts (0.3 ml) of sterile water. The recovery dose of the mix was 0.11-0.13 ml whereas the non-recovery dose was 0.02 ml over and above the recovery dose. The mice were then shaved from the flanks and abdomen so that the mosquitoes could have direct access to the skin. Once the sedation was confirmed,

the mice were then placed on top of the mosquito netted cage containing between 200 and 500 twentyfour-hour starved female mosquitoes. The mosquitoes were then allowed to feed from below onto the abdominal skin of the sedated and secured mice for a period of approximately 20-30 minutes. Once fed on by the infected mosquitoes, the mice were allowed to recover from sedation slowly over a period of 4-6 hours. Blood smears were made from day 3 or 4 post-transmission to ascertain the parasitemia and thus the success of transmission.

Once inside the mouse, the parasite count was continuously monitored through Giemsa-stained smears made from a drop of blood from a tail prick. The parasitemia was allowed to progress up to the desired level as determined by the need of the experiment or until the mouse showed any of the earliest signs of cerebral malaria, whichever was earlier. The infected mouse was then sacrificed through a cardiac puncture (regulated procedure under act) with a sterile 2 ml syringe containing heparin (200 IU/ml in RPMI1640). The animal was then killed by breaking the neck as a part of the regulated procedure under the Act.

### **2.1.1.3 Maintaining infection between mice (weekly mechanical passage)**

Infected mice were serially monitored for parasitemia by Giemsa-stained smear starting from day 5 or 6 of infection, depending on the inoculum. After determining the parasitemia, each week on a fixed day (usually Wednesday), which was day 7 of the infection, the parasites from each of the ten lines were passaged to 10 naïve mice (thus creating 10 new infections each week) after making appropriate dilutions in enriched PBS. For parasitemia of less than 1%, 1 tail-drop of blood was mixed with 1 ml of rich PBS at 4 °C. For parasitemias between 1 and 5% and more than 5%, a 1:10 and 1:100 dilution was done with enriched PBS is done, respectively. Once diluted, 30 µl of the suspension was added to 250 µl of enriched PBS to make up the injection volume and the complete volume (about 280 µl) was injected intra-peritoneally, half on each side of a naïve mouse. The target was to achieve parasitemia between 5 and 10

% on day 7 of weekly passage. When the parasitemia reached this level, the blood from the infected mice was examined under FACS for the presence of gametocytes. Once the presence (or absence) of gametocytes was ascertained by FACS, the mouse was culled by cardiac puncture under general anaesthesia and the blood was cryopreserved for downstream applications. From each mouse, about 1 ml of blood was collected into a 2 ml heparinised syringe (with 26 gauge needle). Half of the blood was mixed with an equal volume of Glycerol/PBS solution (30% Glycerol v/v containing 50 µl Heparin stock (200 IU/ml in RPMI1640) solution), the volume divided into two and immediately cryopreserved in two separate cryotubes in liquid nitrogen. The remaining blood (about 500 µl) was collected in a 1.5 ml microcentrifuge tube at 4°C and centrifuged at 20,238 x g for short duration (8-10 seconds) to separate plasma from the red blood cells. The plasma (usually about 100 µl) was aspirated into a sterile cryotube and immediately cryopreserved in liquid nitrogen.

If a mouse did not show parasites on Giemsa-stained smears even on day 8, a new mouse was infected with one of the cryopreserved stabilates of its immediate predecessor. Each time, one cryotube was withdrawn from the liquid nitrogen store and thawed before injecting the whole volume intra-peritoneally, half on each side of a naïve mouse. This was usually done on a Friday so that the parasites were expected on the next Wednesday, along with the other mice from that week.

The parasites from each of the ten lines were weekly passaged between naïve mice until approximately 52 such passages for each isogenic line were done or when the gametocytes were no longer detected by FACS for a continuous period of 4 weeks, whichever happened earlier.

#### **2.1.1.4 Mosquito handling and maintenance**

All stages of *Anopheles stephensi* mosquitoes, vectors for human and rodent malaria, were reared and maintained in the insectary under the optimum growth conditions of 28 °C temperature and 80% relative humidity. Adult mosquitoes were maintained in netted cages with a holding capacity of approximately 2000 mosquitoes per cage. They were fed through an impregnated

filter paper absorbing 5% fructose from a glass universal bottle placed inside the cage. Once every week, the mosquitoes were blood-fed from 4-week old mice for approximately 10-15 minutes.

Following a blood meal (48-hours post-feed), eggs were collected onto a moistened filter paper placed into an egg bowl. Eggs from adult female mosquitoes were collected and then reared for 48 hours in a hatching tray (filled with low level of water to which 6-8 small fish food pellets were added) until they hatched. Upon hatching, the young larvae were transferred to hatching trays and were again fed on fish food pellets. After about 10 days, the pupae were manually removed from the larval tray by vacuum pump or Pasteur pipette. The pupae were collected into a fresh bowl and kept within a net covered mosquito cage where they metamorphosed to adult mosquitoes.

### **2.1.2      *Parasite growth in vitro***

For experiments requiring the growth of the parasite under specific conditions or drugs or for exclusively growing a specific developmental stage of the parasite, the parasites were maintained out of the mice and or mosquitoes for some period of their life cycle. As *in vitro* maintenance of the full life cycle is currently not possible for *P. berghei*, parasites were ultimately passaged to mice and/or mosquitoes to complete their full life cycle. Freshly harvested asexual stages of the parasites from the mice can be maintained *in vitro* up to the schizont stage (schizont culture) of the parasite, thereafter an intravenous injection of the matured schizonts to the mice is absolutely needed for the continuation of the life cycle as the mature schizonts of *P. berghei* either do not rupture and release merozoites spontaneously or there is very poor re-invasion potential of merozoites in culture (Jambou et al., 2011). Sexual stages of the parasite, i.e., the gametocytes, can be cultured *in vitro* to observe the activation and exflagellation of male gametocytes (exflagellation assay) and the development of ookinetes (ookinete culture) beyond which, the parasite cannot be maintained *in vitro*.

### 2.1.2.1 Schizont culture

For experiments where the pure schizont stage of the parasites was needed, for example, for transfection or for synchronization of infections, approximately 1 ml of infected blood from a phenylhydrazine-treated mouse (at parasitemia between 3% and 4%) was collected through cardiac puncture into a 2 ml heparinized sterile syringe and subsequently transferred into 10 ml of schizont media (RPMI 1640 with L-glutamine and 25 mM HEPES and NaHCO<sub>3</sub> at pH7.3 containing 25% Fetal Calf Serum). The suspension was filtered through CF11 powder (Whatman™) columns (domestically made) equilibrated with schizont media to remove host leukocytes and the filtrate was added to 30 ml of schizont media in a 75 mm<sup>2</sup> tissue culture flask. If the initial parasitemia was around 10% before bleeding, the CF11 filtrate was added to 60 ml of schizont media in a larger tissue culture flask (150 mm<sup>2</sup>). The culture was gassed with a mixture containing 5% CO<sub>2</sub>, 5% O<sub>2</sub> and 90% N<sub>2</sub> for 30 seconds to 1 minute and then incubated in a shaker at 37 °C and 25 rpm overnight. The development of schizonts was assessed next morning (around 9 am) by making a Giemsa-stained blood smear from the schizont culture and examining under a light microscope. When the maturity of schizonts was ascertained by observing the number of merozoites per schizonts, the schizonts were harvested and treated according to desired downstream applications.

### 2.1.2.2 Ookinete culture and count

For assessment of functional maturity of gametocytes, efficiency of fertilization of gametocytes and phenotypic analysis of zygote development, morphology and movement of the ookinetes, ookinete cultures were established from infected blood from mice with asynchronous infections. One or two drops of infected blood were collected from the tail of a phenylhydrazine-treated mouse in approximately 5 ml of ookinete culture medium (RPMI1640, pH 8.00) at 21-22 °C and immediately incubated at 21 °C overnight. Smears were made the next day morning, fixed with methanol and stained with Giemsa to be viewed under 400 x magnifications. Normal ookinetes appear as banana shaped bodies and any deviation from normal appearance indicate an abnormal morphology. Apart from the shape, the number of ookinetes on a slide is an indication of

their adequacy for progression to the next developmental stage, the oocyst. To quantitate the functionality of the ookinetes, another estimate called the ookinete conversion rate was calculated. The ookinete conversion rate reflects the number of female gametocytes/gametes which got converted into ookinetes and is calculated by dividing the total number of ookinetes with the number of female gametocytes/gametes. The number of female gametocytes was determined by counting 30-50 female gametocytes from the Giemsa stained thin blood smears.

### **2.1.2.3 Exflagellation assay**

The process of conversion of male gametocytes into male gametes, called exflagellation, is an important measure of functional maturity of the gametocytes. One or two drops of infected blood were collected from the tail of a phenylhydrazine-treated mouse in approximately 1 ml of the ookinete culture medium. The number of exflagellating male gametocytes in a culture is counted after 10-12 minutes by placing ~10 µl of culture into a Burker cell counter under a light microscope (400 x magnification). The rate of exflagellation was calculated by expressing the total number of exflagellation events per 10<sup>5</sup> erythrocytes.

### **2.1.2.4 Cryopreservation and reactivation**

**Labelling and storage of the stabilate:** Each week, a unique alpha-numeric code was assigned for each of the 10 mice showing successful parasite passage. The same code was applied to the stabilate. The code consisted of a mouse number (m1 to m10; serially denoting the ten isogenic passage lines) followed by a week number (denoting the chronological week of passage, as w1, w2, w3.....). On occasions when the mouse was infected from a predecessor's stabilate (see above), the week number is also followed serially by a lower case letter (for example, a, b, c, d,.....). The letter denoted the number of times a particular line did not come up with parasites and thus had been reactivated in a naïve mouse with infection from a cryopreserved stabilate (for example, "a" means once, "b" means twice, and so on). All the stabilates were stored



together in one rack of a liquid nitrogen tank in a chronological and easily identifiable fashion.

### **2.1.3      *Giemsa staining***

The course of parasitemia (=percentage of infected cells) was determined in Giemsa stained blood films made from tail blood. Briefly, thin blood films from tail blood on a standard microscope slide were made. The smears were air-dried and fixed for 1 second in methanol. Fresh Giemsa solution in 'Giemsa-staining buffer' (10%, v/v) was prepared each time and the smears were dipped in the Giemsa solution to stain for 10 minutes (Fluka Analytical, Sigma). The smears were then rinsed carefully with tap water and the slides were left in an upright position to air-dry. The stained blood films were examined under a standard light microscope with immersion oil and objective at 100 x. Parasites were identified as bodies in stained red blood cells and were counted in fields of uniformly spread out RBCs to avoid overlapping and duplication. At least 10 fields each containing 300-500 red blood cells were examined for parasitemia and the final parasitemia was noted as the percent of red blood cells infected. The life-cycle stage of the parasites was identified by the distinctive morphological features of each stage.

### **2.1.4      *Collection of blood from the infected mouse***

The blood from infected mice was collected by cardiac puncture under Isoflurane anaesthesia (see above) and the animal was sacrificed as per the regulations mentioned in the Animal Licence. A 2 ml heparinised sterile syringe was used to draw approximately at least 0.5 ml of infected blood at the desired parasitemia into an equal volume of glycerol/PBS solution (30% glycerol; v/v), containing 0.05 ml of Heparin stock-solution (200 IU/ml in RPMI1640). The suspension was then transferred to two 2 ml cryotubes, 0.5 ml per cryotube, which were immediately transferred to an appropriate allotted slot in the liquid nitrogen storage tank. When required, the blood suspension from one stabilate (0.5 ml) was used to infect 1-2 mice through an intraperitoneal injection.

### 2.1.5 ***Drugs used during parasite growth and maintenance***

The following drugs were used during parasite growth in mice:

(1) Phenylhydrazine: As *P. berghei* has a strong preference for reticulocytes as compared to mature red blood cells, mice were often treated with phenylhydrazine-HCl (Merck) to induce reticulocytosis. One single dose of 0.1 ml of the (12.5 mg/ml) stock (125 mg/kg body weight) was injected i.p. with a sterile 1 ml syringe 2-5 days prior to the infection of the mice.

Phenylhydrazine was administered to increase the number of asexual blood stages and gametocytes and also for establishment of synchronized parasite infections in mice.

(2) Sulfadiazine: For purification of gametocytes from asynchronous infections, Sulfadiazine (Sigma-Aldrich; 30 mg/L) was dissolved in ddH<sub>2</sub>O and was administered as drinking water to the mice. Sulfadiazine selectively kills the asexual parasite stages while the mature gametocytes remain unaffected. This method yields relatively pure preparations of fully mature (and degenerated) gametocytes. Infected mice were put on the drug when the parasitemia reached ~5% and were subsequently bled after 48 hours of Sulfadiazine treatment.

(3) Pyrimethamine: Stock solution of Pyrimethamine (Sigma-Aldrich; 7 mg/ml) was prepared in DMSO and then diluted 100 times with tap water and the pH adjusted between 3 and 5 using 1M HCl. The drug was administered in drinking water. Pyrimethamine was used as a drug-selectable marker in presence of the *P. berghei* or *T. gondii dhfr* gene which confers resistance to the drug pyrimethamine and hence is useful to positively select for genetically modified parasites bearing the dhfr cassette.

(4) 5-FC: 5-fluorocytosine or 5-FC (Sigma-Aldrich) was used as a selection screen for eliminating the transgenic parasites possessing the negative drug-selectable marker yFCU that combines yeast cytosine deaminase and uridyl phosphoribosyl transferase (UPRT). Parasites containing the *yfcu* gene in their genome develop sensitivity to a lethal metabolite of 5-FC, 5'-fluorocytosine triphosphate (5FCTP) and get killed. The drug was administered either orally in

drinking water (1.5 mg/ml) or as i.p injection of up to 10 mg per day per mouse (from 10 mg/ml stock in 0.9% NaCl), or both.

(5) Hypnovel®/Hypnorm® combination anesthetic: Hypnorm® (Fentanyl-fluanisone; Crown Chemicals) / Hypnovel® (Midazolam; Roche Diagnostics) anesthetic combination approximately 30 minutes prior to mosquito feed. The Hypnorm®/Hypnovel® mix was prepared by adding 1 part (0.15 ml) each of the two drugs to 2 parts (0.3 ml) of sterile water. The recovery dose of the mix was 0.11-0.13 ml whereas the non-recovery dose was 0.02 ml over and above the recovery dose. The cocktail was administered about 30 minutes before the desired procedure, usually a mosquito-bite experiment on the mouse.

(6) Isoflurane: Isoflurane (3-5% in Oxygen) was used as a inhalant general anesthetic for knocking mice down before any surgical procedure.

## **2.2 General molecular biology methods**

### **2.2.1 *Isolation of genomic DNA***

In order to analyse the genotype of the transfected or cloned parasites or for performing broader downstream applications including PCR based analyses, restriction digest, molecular cloning, etc, parasite genomic DNA had to be extracted from the infected mouse blood. For this, infected blood, usually through a cardiac puncture, was collected in 10 ml of PBS. The suspension was filtered through the CF11 cellulose powder columns or Plasmodipur Filters (EuroProxima B.V., The Netherlands) to remove host leukocytes. The filtrate was spun to obtain the pellet which was dissolved in 40 ml of chilled 1x RBC lysis buffer. Lysis was allowed to happen over the next 10-20 minutes on ice or until the suspension became semi-transparent. The suspension was spun again to obtain the parasite pellet which was washed twice with PBS to remove the supernatant. The final PBS washed parasite pellet was then used to extract the DNA.

The pellet in a microcentrifuge tube was resuspended in 700 µl TNE buffer and the following supplements were added - 200 µg RNase A (Invitrogen), 1% (v/v) SDS, and demineralised water up to 1 ml. The lysate was incubated for 10 minutes at 37 °C after which 200 µg of Proteinase K was added, with incubation for 1 hour at 37 °C. Buffered phenol was then added up to 1.5 ml and the tube inverted several times and centrifuged for 5 minutes at 20,238 x g. The aqueous upper phase was transferred to a new tube and buffered phenol:chloroform:isoamylalcohol (25:24:1) was added up to 1.5 ml. The tube was again inverted several times and centrifuged for 5 minutes at 20,238 x g. The aqueous upper phase was again transferred to a new tube and this time chloroform:isoamylalcohol (24:1) was added up to 1.5 ml. The tube was again inverted several times and centrifuged for 5 minutes at 20,238 x g. The aqueous upper phase was again transferred to a new tube and 0.1 volume of 3 M NaAc, pH 5.2, and 2 volumes of 96% ethanol were added. The reactions were gently mixed and the DNA was allowed to precipitate at -20 °C overnight. The next morning, the precipitate was centrifuged for 10 minutes at 20,238 x g at 4 °C and the DNA pellet washed by adding 500 µl 70% ethanol and centrifuged 5 minutes at 20,238 x g at 4 °C. The supernatant was discarded and the pellet air-dried. The DNA was re-suspended in 50 µl of demineralised water.

### **2.2.2      *Polymerase Chain Reaction (PCR)***

PCR was used to amplify specific regions of genomic and complementary DNA that would be a template for the following downstream applications: Sanger sequencing, molecular cloning, to confirm integration in transfected cell lines, PCR-based complementation attempts, insertion of a restriction site, etc. All oligonucleotides were synthesised by Eurofins MWG Operon (Ebersberg, Germany). The standard PCR reaction used was as follows; 1x PCR reaction buffer without MgCl<sub>2</sub> (Invitrogen, Paisley, UK), 1.5 mM MgCl<sub>2</sub> (Invitrogen, Paisley, UK), 0.4 µM of each primer (Eurofins MWG Operon, Ebersberg, Germany), 0.2 mM dNTPs (Invitrogen, Paisley, Scotland), 1.0 U Taq Polymerase (Invitrogen, Paisley, UK), and between 1 and 5 ng of plasmid DNA, or between 50 and 100 ng of gDNA template. The amplification was performed using a thermal cycler (PTC-200 DNA Engine Thermal Cycler, BioRad). The PCR program usually included an initial

denaturation step at 94 °C for 3 minutes was followed by between 25 and 35 cycles of 94 °C for 30 seconds, an annealing temperature ( $T_m$ ) specific for each set of primers for 30 seconds, and an extension step of 1 minute per kb of the expected length of the amplified DNA. A final extension step at 68 °C (or 72 °C for amplicons less than a kb) for 10 minutes was followed by a holding temperature of 4 °C.

For the amplification of DNA for sequencing, an error-proof DNA polymerase system, Expand High Fidelity PCR system (Roche Diagnostics, Mannheim, Germany) was used. The reaction components used was as follows; 1x Reaction Buffer including 1.5 mM  $MgCl_2$ , 200  $\mu M$  of each dNTP, 300 nM of each primer, 50-100 ng of template gDNA, 2.6 U of High Fidelity enzyme mix for each 25  $\mu l$  reaction volume. The cycling conditions were as follows; an initial denaturation step at 94 °C for 2 minutes was followed by 10 cycles of 94 °C for 15 seconds, an annealing temperature depending upon the  $T_m$  of the primers for 15 seconds, and an extension temperature of 68/72 °C for time depending on the fragment length (45 seconds for up to 0.75 kb, 1 minute for 1.5 kb, 2 minutes for 3 kb, 4 minutes for 6 kb, 8 minutes for 10 kb). A final extension step at 68/72 °C for 7 minutes was followed by a holding temperature of 4 °C.

### **2.2.3      *Oligonucleotides***

All oligonucleotides were designed using Vector NTI (Invitrogen, Paisley, Scotland) and CLC Genomics Workbench 5.0 software and ordered from Eurofins MWG Operon (Ebersberg, Germany). A tabulated description of all the oligonucleotides used in the study is mentioned in Table A-13 (Appendices).

### **2.2.4      *Southern analysis***

Approximately 10  $\mu g$  of purified genomic DNA was digested with appropriate restriction enzymes (New England BioLabs, Massachusetts). A typical double digest consisted of 150  $\mu l$  of reaction mix with not more than 10% of the total reaction volume as both the restriction enzymes (a total of 15  $\mu l$  of both enzymes) in presence of BSA (if recommended) and appropriate NEB reaction

buffer. The reaction was incubated for ~4 hours at 37 °C (usually or as recommended) after which additional 2% of the total reaction volume of each enzyme (that is 3 µl of each enzyme) was supplemented to the reaction and digestion was continued at 37 °C for another 1 hour. The reaction mix was vacuum concentrated (SpeedVac; Thermo Scientific) to ~35 µl. The concentrated samples (together with 5x loading dye) were loaded on a 0.7% agarose gel (for fragments of 0.8 kb to 10 kb) in water (added SYBR®Safe DNA gel stain, Life Technologies™) and fractionated at 20 Volts/cm overnight with circulating 1x TBE buffer. When the gel has run more than 3/4<sup>th</sup> of its length - a UV picture of the gel with a ruler was taken and ladder sizes marked on the picture. The DNA was then depurinated with 0.25 M HCl and incubated on a slow shaker for 15 minutes at room temperature (RT). Denaturation of the DNA was done by rinsing the gel with ddH<sub>2</sub>O and covering with denaturing buffer (see appendix A14) and incubated on a slow shaker for 20-30 minutes at RT. The DNA was neutralized by soaking the gel in 20x SSC for 30 minutes with constant shaking at RT.

Rest of the steps including the transfer of the DNA from the gel, pre hybridization and hybridization steps, washing and visualization of bands were the same as that for RNA/northern analyses (see below).

### **2.2.5      *Isolation of total RNA***

High quality total RNA was usually obtained from “fresh” parasite pellets obtained directly from the animal or from overnight *in vitro* culture. The procedure for parasite RNA extraction is same as for extraction of DNA up to the stage of obtaining the parasite pellet. Thereafter, the parasite pellet was resuspended in 1 ml of TRIzol® reagent (Invitrogen, Life Technologies) and vigorously vortexed until a homogenous solution was obtained. The solution was pelleted and the supernatant was transferred to a new microcentrifuge tube. Two hundred µl of chloroform was added and the sample vortexed for 30 seconds and then incubated for 10 minutes at room temperature. The sample was then centrifuged at 20,238 x g , at 4 °C for 10 minutes in order to obtain phase separation. The upper aqueous phase was transferred to a fresh

microcentrifuge tube and 500 µl of isopropanol added. Precipitation was performed at -20 °C overnight. On occasion 20 µg of glycogen was added as a co-precipitant in order to enhance precipitation. The RNA was pelleted by centrifugation at 20,238 x g for 15 minutes at 4 °C, and the supernatant decanted. The pellet was washed with 1 ml of 70% ethanol and spun again for 10 minutes, at 20,238 x g at 4°C. The supernatant was removed and the pellet dried for up to 10 minutes at room temperature. The dry pellet was then resuspended in 30-50 µl RNase-free demineralised water (depending on the pellet size). The RNA concentration was measured using a NanoDrop 2000 spectrophotometer (Thermo Scientific). RNA samples were stored at - 80 °C.

### **2.2.6      *Preparation of cDNA and RT-PCR***

All RNA processing was done in an RNase free environment by profound use of Ambion® RNaseZap® (Life Technologies™). Purified total RNA (~ 4 µg) was mixed with 10x DNase I reaction buffer (Invitrogen) and 2 U DNase I amplification grade (Invitrogen) in DEPC treated demineralized water in RNase free microcentrifuge tubes. The reaction was incubated at room temperature for 15 minutes after which the DNase I was inactivated by adding 1 µl of 25 mM EDTA and heating for 10 minutes at 65 °C. First strand cDNA synthesis was done using SuperScript III RT (Invitrogen). The reaction mix consisted of the following in a final volume of 20 µl- ~5 µg of total RNA (from above), 50 µM oligo (dT) 20, and 10 mM each of the dNTPs plus sterile distilled water. The reaction was incubated at 65 °C for 5 minutes followed by on ice for a minute. The reaction was centrifuged at 20,238 x g and the following added - 5x first strand buffer, 0.1 M DTT, 40 U RNaseOUT (Invitrogen) and 400 U of SuperScript III RT (Invitrogen). The reaction was mixed gently and incubated at 65 °C for 30-60 minutes. The reaction was terminated by incubating at 70 °C for another 15 minutes. To remove RNA complementary to the cDNA, 1 µl (2 units) of *E. coli* RNase H (Invitrogen) was added and samples incubated at 37 °C for 20 minutes. The first strand cDNA thus synthesised was used as a template for PCR, with a positive gDNA control and negative water and minus-RT controls.

### 2.2.7 *Northern analysis*

Freshly prepared and purified total RNA from the schizont preparation of *Plasmodium* parasites was used for northern analyses. Purified total RNA and an RNA ladder (Life Technologies<sup>TM</sup>; 1 µg/µl) were diluted in DEPC treated water to obtain ~5 µg of RNA sample and 7 µg of the RNA ladder each in 12 µl total volume. RNA was then denatured by adding 12 µl of 2X RNA loading dye (Thermo Scientific) to each 12 µl of RNA sample and heating at 70 °C for 10 minutes followed by chilling on ice for 5 minutes. The samples were loaded on a 1.2% agarose gel (with added SYBR®Safe gel stain, Life Technologies<sup>TM</sup>) in DEPC water (containing 10x MOPS and 37% Formaldehyde). The RNA was fractionated at 20 Volts/cm overnight in 1x MOPS running buffer.

Next morning, the gel was soaked in DEPC water to remove formaldehyde (fluoresce under UV) and then photographed under UV light alongside a ruler to mark the ladder, wells and detect the 4 distinct bands of parasite rRNA. The RNA was then transferred from the gel onto a nylon membrane (Hybond-N+<sup>TM</sup> nylon transfer membrane, GE Healthcare) pre-wetted in 2x SSC using a TurboBlotter<sup>TM</sup> device (Whatman<sup>TM</sup>, GE Healthcare). Briefly, the gel (as is) was placed on the pre-wetted nylon membrane with a stack of SSC-soaked filter papers on the gel as weights. The stack further supported a pre-wetted filter paper wick on top which was immersed in the 10x SSC transfer buffer. A light cover plate was just placed on the wick but not pressed. The gravity-aided capillary transfer of RNA from gel to the membrane was allowed to occur overnight.

Next morning, transfer efficiency was assessed by inspecting the membrane under UV and subsequently crosslinking the RNA to the membrane using ultraviolet light. The membrane was first incubated with pre-hybridization buffer at 55 °C for an hour before hybridization with denatured  $\alpha$ -<sup>32</sup>P dATP labelled double stranded probe specific to the analysed sequence at 60 °C overnight with gentle shaking. After overnight hybridisation, the blot was washed three times with 3x SSC/0.5% (v/v) SDS and once with 1x SSC/0.5% (v/v) SDS at 60 °C (each washing step for 15 minutes). The signal was detected using a storage phosphor screen (Kodak).



### **2.2.8      *Generation of gene knockout constructs***

The plasmid pL0034 (Figure 2-1) was used to generate all of the three length variants of the gene knockout constructs. The plasmid pL0034 was selected because of the presence of only one copy of the 3'Pbdhfr UTR which would favour negative selection-based complementation of the knocked out gene. Conventional double crossover gene targeting constructs were developed by amplifying approximately 1 kb upstream and downstream homology arms flanking the region of the gene to be knocked out. The oligonucleotides used for amplifying each of these regions were synthesised by Eurofins MWG operon (Ebersberg, Germany) and were so designed to incorporate the desired restriction sites necessary to clone the homology arms into the pL0034 vectors. The homology arms (inserts) and the vector were digested with compatible restriction enzymes and the inserts sequentially cloned into the vector to generate the complete gene targeting vector bearing the homology arms flanking the drug selectable marker cassette. The recombineering-based gene targeting vector was kindly supplied by Oliver Billker, Sanger Institute, Cambridge, UK.

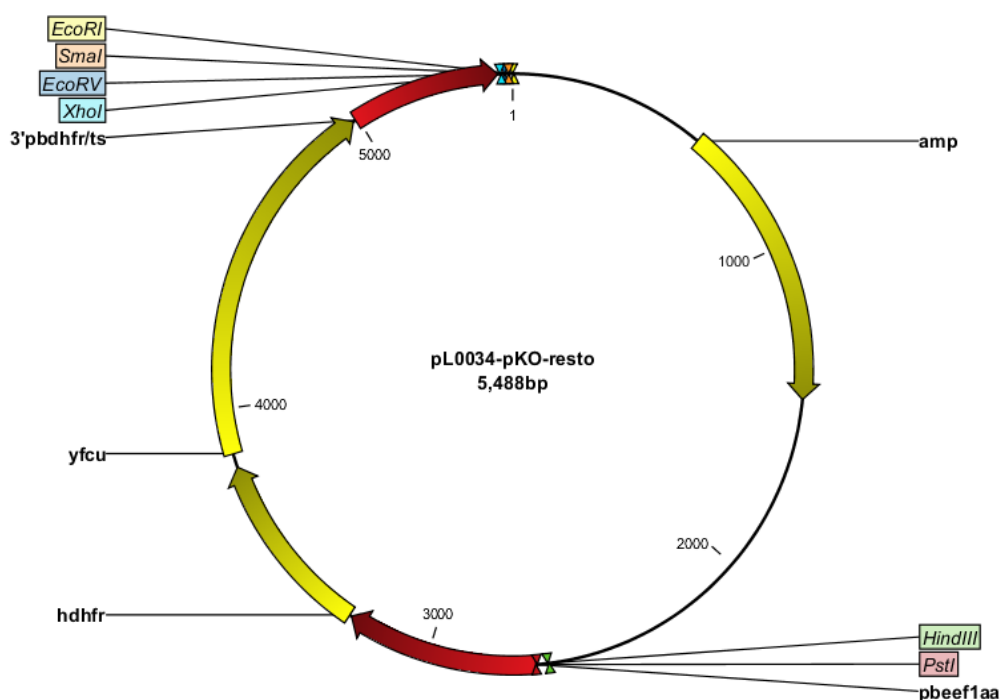
### **2.2.9      *Generating gene complementation constructs***

Two types of complementation constructs were used to restore the function of the knocked out gene - PCR-based constructs and recombineering based constructs. PCR based constructs were synthesised by conventional PCRs encompassing the region of the gene to be complemented using a “proofreading” polymerase system, the Expand High Fidelity PCR system (Roche Diagnostics, Mannheim, Germany). Recombineering-based complementation constructs were kindly supplied by Oliver Billker, Sanger Institute, Cambridge, UK.

## 2.2.10 Parasite transfection

### 2.2.10.1 Preparation of DNA for transfection

DNA constructs for transfection were prepared according to standard molecular biology techniques. The homology arms for all generated constructs were obtained by PCR on wild type gDNA with standard PCR program conditions (see above). The amplified DNA was quantified using a spectrophotometer and an aliquot (usually 50-100 ng) of the obtained PCR products fractionated on 1% agarose gel to confirm the desired size of the homology arm/s. The remaining PCR product was purified using QIAquick PCR purification kit (Qiagen) as per the manufacturer's instructions. The purified DNA was again quantitated as above.



**Figure 2-1. A cartoon of the pL0034 plasmid (size 5.4 kb) used for generating the vectors for targeting the *AP2-G* gene.** Note that the plasmid has only one copy of the 3'pbdhfr/ts which prevents the usage of this particular vector for selectable marker recycling. Also, for the same reason, the vector is ideal for any negative selection based enrichment experiment, as in the case here, the complementation of the gene at the endogenous locus. The restriction sites at the multiple cloning sites are also depicted.

For direct cloning into the desired plasmid vector, each PCR product or homology arm (the insert) was digested with appropriate restriction enzymes for approximately 1 hour at 37°C (or as specified for the target restriction enzymes) and again purified with QIAquick PCR purification kit (Qiagen, Manchester, UK).

The target plasmid was also simultaneously digested with corresponding appropriate restriction enzymes (double digestion) and purified as above (the vector). The concentration of both the insert and the vector was determined by use of the spectrophotometer.

Vector and the insert DNA were ligated in a molar ratios ranging from 1:1, 1:3 and 1:6. The ligation reaction was performed for 10 minutes at room temperature using Rapid DNA ligation kit (Roche Applied Science, Mannheim, Germany). A typical 15 µl ligation reaction consisted of the vector and insert at chosen molar ratio, 1.5 µl of 10x ligation buffer (Roche Applied Science, Mannheim, Germany) and 1 U of T4 DNA ligase (Roche Applied Science, Mannheim, Germany). The ligation mix was then added to 50 µl of ice-thawed Fusion-Blue competent *E.coli* cells (Clonetech), and the mix incubated on ice for 30 minutes. Heat-shock transformation was performed at 42 °C for 45 seconds, followed immediately by incubation on ice for 1 minute. About 400 µl of room temperature SOC or LB medium was then added to each transformation reaction and incubated with shaking at 37 °C for an hour. The transformed bacteria were then spread on pre-warmed LB agar plate with 100 µg/ml ampicillin and incubated overnight at 37 °C. The next day bacterial colonies were inoculated in 4 ml LB medium with 100 µg/ml ampicillin. After overnight incubation, the cultures were spun down and the plasmid DNA extracted using the QIAprep Spin Miniprep kit (Qiagen, Manchester, UK) as per the manufacturer's instructions. A sample of the obtained plasmids were analysed by appropriate restriction enzyme digestion to confirm the ligation.

Approximately 20 µg of the final plasmids with introduced homology arms were linearized with suitable restriction enzymes overnight at 2 µg per reaction with 10U/reaction of each of the enzymes in the appropriate buffer, at the appropriate temperature to ensure complete digestion. For the double crossover constructs, the plasmid was linearized using restriction sites at the end of the desired sequence target construct to separate the plasmid backbone from the desired linear recombination construct. The plasmid was also digested simultaneously with another unique restriction enzyme to cut the backbone to generate an easily separable desired linearized target construct and ensure minimal contamination with the circular construct.

The reactions were pooled and the digested plasmid (along with the undigested plasmid as a control) fractionated usually on an 0.8% agarose gel containing a minimal amount of SYBR®safe (Invitrogen, Paisley, UK) to differentiate uncut and the linearized plasmid. The band corresponding to the linearized plasmid was cut out the gel (with standard precautions to prevent contamination) as quickly as possible to minimize exposure to UV rays. DNA was extracted from the gel using QIAquick Gel Extraction kit (Qiagen, Manchester, UK) and precipitated with 0.1 volume of 3M NaAc, pH 5.2, and 2.5 volumes of 96% ethanol overnight at -20 °C. Next morning, the DNA was spun down at 20,238 x g for 15 minutes in a centrifuge at 4 °C, washed with 200 µl of 70% ethanol and spun down again at 20,238 x g for 10 minutes at 4 °C. The DNA was air-dried and resuspended in 5-10 µl of double-deionized water so as to make a final concentration of around 1 µg/µl. For transfection of *P. berghei* with the Amaxa® Human T-cell Nucleofector™ solution-buffer (Lonza, Germany), 5 µg of DNA was used for each transfection.

TOPO® TA cloning® (Invitrogen, Life Technologies) was done as an easy and fast way to ligate PCR products directly into the TOPO vector, thus avoiding the step of enzymatically making the insert compatible with the vector. The technology is based on the property of the Taq polymerase to add extra adenosines at the 3' end of the amplified products which remain compatible to the complementary T sequences in the TOPO® vector. The ligations were done using the TOPO® TA cloning® kit for sequencing (Invitrogen, Life Technologies) as per the manufacturer's instructions. Basically, 0.5-2 µl of fresh PCR product was incubated for 5 minutes at room temperature with 0.5 µl of TOPO® vector and sterile water in a 3 µl total reaction volume. The reaction was incubated on ice while a vial with One Shot® Top 10 Chemically Competent *E. Coli* cells (Life Technologies, Paisley, UK) was thawed on ice. Two µl of the cloning reaction was added to the vial of competent cells and mixed gently. After incubation on ice for 5-30 minutes, the cells were heat-shocked for 30 seconds at 42 °C without shaking and were immediately transferred to ice. The cells were then incubated with 250 µl of room temperature SOC medium at 37 °C for 1 hour with shaking. Finally, 10-250 µl of the reaction was spread on a LB-ampicillin plate and incubated at 37 °C overnight.

### 2.2.10.2 Preparation of schizonts for transfection

Matured *P. berghei* schizonts were used to transfect the linearized plasmid targeted construct. A phenylhydrazine-treated mouse was infected with the cryo-preserved parasites of the line to be genetically manipulated to serve as a source of blood stage parasites for the culture and purification of schizonts. The parasites were harvested on day 4 or 5 and put into schizont culture (see above). Prior to the electroporation, the schizonts were separated from the uninfected erythrocytes using 55% Nycodenz (Life Technologies, Paisley, UK)/PBS density gradient solution (v/v). Briefly, the culture suspension containing the schizonts was transferred to four 50 ml tubes (35 ml per tube). Using a 10 ml pipette, 10 ml of the Nycodenz-solution was gently added in each tube under the culture suspension, so that a sharp contrasting division is visible between the two suspensions. The suspension was then centrifuged for 20-30 minutes at 1500 rpm with no brake. The brown/grey layer (containing schizont infected cells) at the interface between the two suspensions was carefully collected. The schizonts were then pelleted and washed with culture medium obtained from the top of the gradients. The supernatant was discarded and 100 µl of the Amaxa® Human T-cell Nucleofector™ solution-buffer (Lonza, Germany) containing 5-10 µl of DNA (5-10 µg DNA construct in water or TE buffer) was added to resuspend the parasites. DNA transfection was then performed with the Nucleofector™ 2b Device (Lonza, Germany) using protocol U33. Immediately after transfection, 50 µl of the culture medium was added and the transfected parasites were immediately injected (i.v.) into mice. Each transfection was allocated a unique alphanumeric serial number for identification like G418.

### 2.2.11 Cloning of transgenic parasites by limiting dilution

A transfected parasite population normally consists of both the transfected (transgenic) and the wild type parasites. Therefore cloning by limiting dilution is necessary for further analysis of a homogenous transgenic line. Usually, on day 0, a phenylhydrazine-treated donor mouse was infected intraperitoneally with 100 µl of a blood-suspension from cryo-preserved transfected parasites of the line to be cloned. On day 3 or 4 parasitemia was

checked (usually the parasitemia ranged between 0.3 and 1%) and 5  $\mu$ l of tail blood was collected by a heparin-treated capillary tube into a microcentrifuge tube. The blood was then diluted in 1 ml of enriched PBS and 10  $\mu$ l was used for red blood cell counting using a Bürker cell counter. The number of uninfected and infected erythrocytes per  $\mu$ l of the cell suspension was calculated and the sample diluted to a final concentration of 0.5 parasites/ $\mu$ l of enriched PBS solution. This solution was further diluted 200 times and 200  $\mu$ l of the final diluted suspension was injected per mouse intravenously such that less than 1 parasite is injected per mouse (in total 10 mice were used per cloning).

Thus, 10 mice were infected for obtaining a cloned parasite line. According to the calculations above, 0.5 parasites are injected intravenously (i.v.) into each mouse, resulting in an infection rate of 20-50% of the mice (due to statistical probability, not every mouse receives a parasite). At day 8 after infection the parasitemia was determined. From the mice with a parasitemia of 0.3 - 1%, approximately 1 ml of blood was collected through cardiac puncture. Half of the collected blood volume was used for extraction of DNA (for PCR based confirmation of the success of cloning experiment) and the remaining half was used to make stabulates for cryopreservation of the anticipated cloned line.

## 2.3 Recombinant protein production

N-terminal glutathione S-transferase (GST) fused extended ApiAP2 DNA binding domains (cloned into pGEX-4T1) from *P. falciparum* AP2-G (PFL1085w) and *P. berghei* AP2-G (PBANKA\_143750) were kindly gifted by Manuel Llinas (Princeton University, USA) as plasmid DNA. Rosetta™ (DE3)pLysS competent cells (Merck Millipore, Germany) were transformed with the provided plasmid DNA (as above) and plated on LB/ampicillin (amp)/Chloramphenicol (cam) plates.

Starter cultures were prepared by picking one colony from an LB/amp-cam-plate into 10 ml LB containing 10  $\mu$ l ampicillin (100 mg/ml stock) & 10  $\mu$ l chloramphenicol (35 mg/ml stock). The cultures were grown overnight at 37 °C with shaking. Next morning, 100 ml of fresh LB/amp/cam was inoculated with 2

ml of overnight culture. The main culture was grown again at 37 °C for ~1 hour (up to an OD600 of ~0.2). The cultures were then transferred to 25 °C and grown for ~2 hours (up to an OD600 of ~0.6). Expression of the GST-tagged protein was induced with 0.2 mM of fresh UltraPure™ IPTG (Invitrogen, Life Technologies) and the culture was further incubated at 25 °C for 4 more hours. After that, the culture was spun at 6000 x g for 10 minutes to pellet cells. LB medium was decanted and the pellet was stored at -80 °C. The successful induction of the desired protein expression was ascertained by fractionating the proteins in a 12% SDS PAGE gel to determine if there is any protein in the induced sample that is significantly more than that in the un-induced one.

The protein pellet was re-suspended in freshly prepared GST buffer at pH 8 and cell lysis achieved using a cocktail of the following - 1 mM PMSF (Sigma-Aldrich), 0.5% Triton X-100 detergent (Thermo Scientific), 1 mg/ml Lysozyme (Sigma-Aldrich), 100 U/ml of DNase (Invitrogen, Life Technologies) and 1 tablet of protease inhibitor (cOmplete, Mini, EDTA free protease inhibitor cocktail tablets, Roche). The suspension was incubated for 30 minutes at 4 °C on a nutator at 20 rpm. To ensure completeness of bacterial cell lysis, the suspension was freeze-thawed and sonicated (thrice) in ice bucket in short 10-second bursts alternated with 1 minute of resting on ice. The completeness of lysis was indicated when the cloudy cell suspension became translucent. The cell debris was spun down immediately at 20,238 x g for 10 minutes at 4 °C. The soluble proteins in the resulting supernatant were batch purified using Glutathione resin (Glutathione HiCap Matrix slurry; Qiagen). Briefly, Glutathione beads were equilibrated and prepared as per manufacturer's instructions and the cell lysate (supernatant) added to the equilibrated Glutathione beads. The suspension was then incubated at 4 °C on the nutator for 4 hours at 20 rpm. The resin was allowed to settle for 5 minutes and the supernatant was drawn off. The resin was then washed 3 times with cold GST buffer, allowed to settle for 5 minutes and the supernatant drawn off. Finally, the GST-tagged protein was eluted using freshly prepared elution buffer at pH ~8.0 (containing reduced L-Glutathione (Sigma-Aldrich), Tris-HCl, ddH<sub>2</sub>O and protease inhibitor tablets (cOmplete, Mini, EDTA-free protease inhibitor cocktail tablets, Roche)). Elution was repeated thrice, with increasing duration of incubation, using 500 µl of the elution buffer each time in a nutator at 20 rpm. Each time, the resin was allowed to settle by

gravity and supernatant (eluate) containing the pure protein drawn off carefully. The purified proteins (elutions) were stored at 4°C. To render the tagged proteins directly usable in EMSAs, buffer exchange was performed with 1X PBS and proteins concentrated at the same time using Amicon-Ultra 30kDa spin concentrators (Amicon Ultra-0.5 Centrifugal filter devices; 30K device; Milipore). For long term storage, purified proteins were stored as pellets at a concentration 1 mg/ml in -80°C in 25% glycerol, aliquoted into 100 µl PCR tubes.

## 2.4 Electro-mobility Shift Assays (EMSA)

Binding of purified N-terminal GST fusions of AP2 domains of AP2-G of *P. falciparum* (PF3D7\_1222600) and *P. berghei* (PABNKA\_143750) to their cognate DNA sequences was analysed using EMSAs. Single stranded oligonucleotides containing the DNA-binding Domain (DBD) recognition motif flanked both by random nucleotides (same for all flanking sequences) and by actual nucleotides (as they occur naturally in the 5' upstream regions of potential AP2 target genes) and their corresponding complementary oligonucleotides were purchased from Eurofins MWG Operon (Ebersberg, Germany) as labelled (5'-biotinylated & HPLC purified) and unlabelled sequences. These single-stranded oligonucleotides were annealed to their complementary oligonucleotides to create double-stranded probes by heating the oligonucleotides, in equi-molar concentration, at 95°C for 3-5 minutes followed by slow cooling at room temperature for about an hour.

These annealed oligonucleotides were used as labelled and unlabelled target probes for the DNA-binding domain of AP2-G in non-competitive and competitive EMSAs, respectively EMSAs were done using the LightShift Chemiluminescent EMSA kit (Pierce Protein Biology Products, Thermo Fisher Scientific, Northumberland, UK). For non-competitive EMSA, 2 µg of the purified GST fusion of PFL1085w and PBANKA\_143750 (in separate reactions) was incubated with the labelled probe (0.02 pmol) in 20 µl of the binding reaction containing 10X binding buffer, 1 µg poly(dI-dC), 50% glycerol, 100 mM MgCl<sub>2</sub>, 1% NP40, and 60 µg BSA at room temperature for 20 minutes in 1.5 ml microcentrifuge tubes, protected from light. For competitive EMSA, the same amount of respective fusion protein was first pre-incubated with the unlabelled



probe (4 pmol; 200-fold excess to the labelled probe) in 20 µl of the binding reaction as above. The labelled probe (0.02 pmol) was then added to the reaction and incubated for further 20 minutes at room temperature.

In either case, the reaction was fractionated using 12% polyacrylamide gel (without SDS) in 0.5x TBE at 100 Volts until the dye had run  $2/3^{\text{rd}}$  -  $3/4^{\text{th}}$  of the gel length. The nucleic acid (bound and unbound) was then transferred from the gel to a nitrocellulose membrane (Amersham Hybond Nitrocellulose membrane, GE Healthcare, Buckinghamshire, UK) as per manufacturer's instructions. Briefly, the transfer cassette was assembled by placing the gel on to the Nitrocellulose membrane such that the dye-front of the gel faced the membrane and the two were sandwiched between two pieces of wetted Whatman™ paper (GE Healthcare) and the transfer sponges (reserved specifically for EMSAs). The transfer was performed at 100 Volts for 30 minutes in chilled 0.5x TBE buffer under cold room conditions.

The membrane was then UV-cross-linked. Specific binding of the AP2 domain with the target motif was detected as an upward shift using the Chemiluminescence Nucleic Acid Detection Module (Pierce Protein Biology Products, Thermo Fisher Scientific, Northumberland, UK) as per the manufacturer's instructions. The membrane was first blocked with a blocking buffer for 15 minutes shaking at RT and then conjugated with Stabilized Streptavidin-Horseradish Peroxidase Conjugate for another 15 minutes. The membrane was then washed four times with 1x wash buffer for 5 minutes each and subsequently equilibrated with Substrate Equilibration Buffer for 5 minutes with shaking. The membrane was placed on clean saran wrap and then incubated with Luminol/enhancer and stable peroxide solution for a further 5 minutes in the dark without shaking. Chemiluminescence was detected using a storage phosphor screen with exposure for 30 seconds to 5 minutes on Kodak™ BioMax® MS (maximum sensitivity) and MR (maximum resolution) films.

## 2.5 Flow cytometry

Each week, gametocytogenesis in each of the infected lines was inferred from their expression of green and red fluorescent proteins and DNA content (or

ploidy) determined by FACS measurement of fluorescence intensity of cells stained with the DNA-specific dye Hoechst-33258 (Sigma-Aldrich) using a MACSQuant® Analyzer (Flowcytometer; Miltenyi Biotec, Germany). GNP lines were detected by the absence of gametocytes which the FACS analyses confirmed.

### **2.5.1 *Hoechst staining of the samples***

In short, 2-3 drops of tail blood was collected in 1 ml of enriched PBS at 4 °C in a dark-coloured microcentrifuge tube kept on ice. The suspension was spun at 20,238 x g for 2 minutes to pellet the blood cells. The pellet was washed twice with enriched PBS before being finally stained with 1 ml of 0.2 µM Hoechst stain (Sigma-Aldrich) in enriched PBS. The suspension was always kept on ice to prevent the activation of male gametocytes. Once Hoechst was added, the suspension was incubated for 1 hour at 37 °C. Following incubation, the cells were examined using the MACSQuant® Analyzer (Flowcytometer; Miltenyi Biotec, Germany).

### **2.5.2 *Instrument settings for MACSQuant® Analyzer***

Each sample was run on the flowcytometer with a sample volume of 1000 µl, uptake volume of 200 µl, on a standard mode with customized settings. A total of 200,000 events were counted with an average flow rate of up to 15000 events per second.

Five optical parameters of cells were analysed - three fluorescence channels and two scatter channels (one forward scatter or FSC and one side scatter or SSC channel). The cells were excited with a wavelength of 405 nm for FL1 (VioBlue/Hoechst) channel and 488 nm for FL2 (fluorescein isothiocyanate/GFP or FITC), FL3 (Phycoerythrin/RFP or PE) and FSC/SSC channels. Cells expressing GFP have an emission maximum of 521 nm and were hence detected in the green fluorescent channel (FL2/FITC, 525 nm with a bandwidth of 50 nm). Cells expressing RFP have an emission maximum of 585 nm and were hence detected in the red fluorescent channel (FL3/PE, 585 nm with a bandwidth of 40 nm). Thus, the following filters were used - 450/50 for FL1/VioBlue, 525/50 for FL2/FITC, 485/40 for FL3/PE and 488/10 for FSC/SSC.

A linear scale was used for FSC and SSC data whereas a logistic scale was used to analyze fluorescence signals as a difference in fluorescence signal intensities extended over several orders of magnitude. Cells for Hoechst analysis were selected on size by gating on FSC and SSC. A sample volume of 1000  $\mu$ l of Hoechst-stained cells with an uptake volume of 200  $\mu$ l was examined using 200,000 cells per sample run. Data processing and analysis was performed using the MACSQuantify™ analysis software ([www.miltenyibiotec.com](http://www.miltenyibiotec.com)). Dot plot, density plots and histograms were used to analyze the distribution of cells.

Cells on FSC versus SSC dot plot are widely gated (P1) to include most of the red blood cells while eliminating larger (lymphocytes and monocytes) and granular (granulocytes) leukocytes and smaller cellular debris. The gated population of cells (P1) was analyzed further on FITC versus VioBlue density plot, PE versus VioBlue density plot and FITC versus PE density plots. The FITC versus VioBlue density plot differentiated red blood cells that were only Hoechst positive (infected red blood cells; asexuals) from red blood cells that were positive for both Hoechst and GFP signals (infected red blood cells; male gametocytes). A second gate (P2) captured the male gametocytes in this plot and estimated the absolute number and proportion (as %) of male gametocytes out of the total population of red blood cells gated (P1) in FSC versus SSC dot plot.

Similarly, the PE versus VioBlue density plot differentiated red blood cells that were only Hoechst positive (infected red blood cells; asexuals) from red blood cells that were positive for both Hoechst and RFP signals (infected red blood cells; female gametocytes). A third gate (P3) captured the female gametocytes in this plot and estimated the absolute number and proportion (as %) of female gametocytes out of the total population of red blood cells gated (P1) in FSC versus SSC dot plot. A histogram of VioBlue positive cells from the P1 gate estimated the proportion of red blood cells infected out of the total population of red blood cells gated as P1. This showed the parasitemia as percent of red cells infected. A red cell population that was multiply infected was apparent as a separate density in the VioBlue density plots. The FITC versus PE density plot revealed the comparative presence of male (P2 gated) and female (P3 gated) gametocytes, respectively. The activated male gametocytes

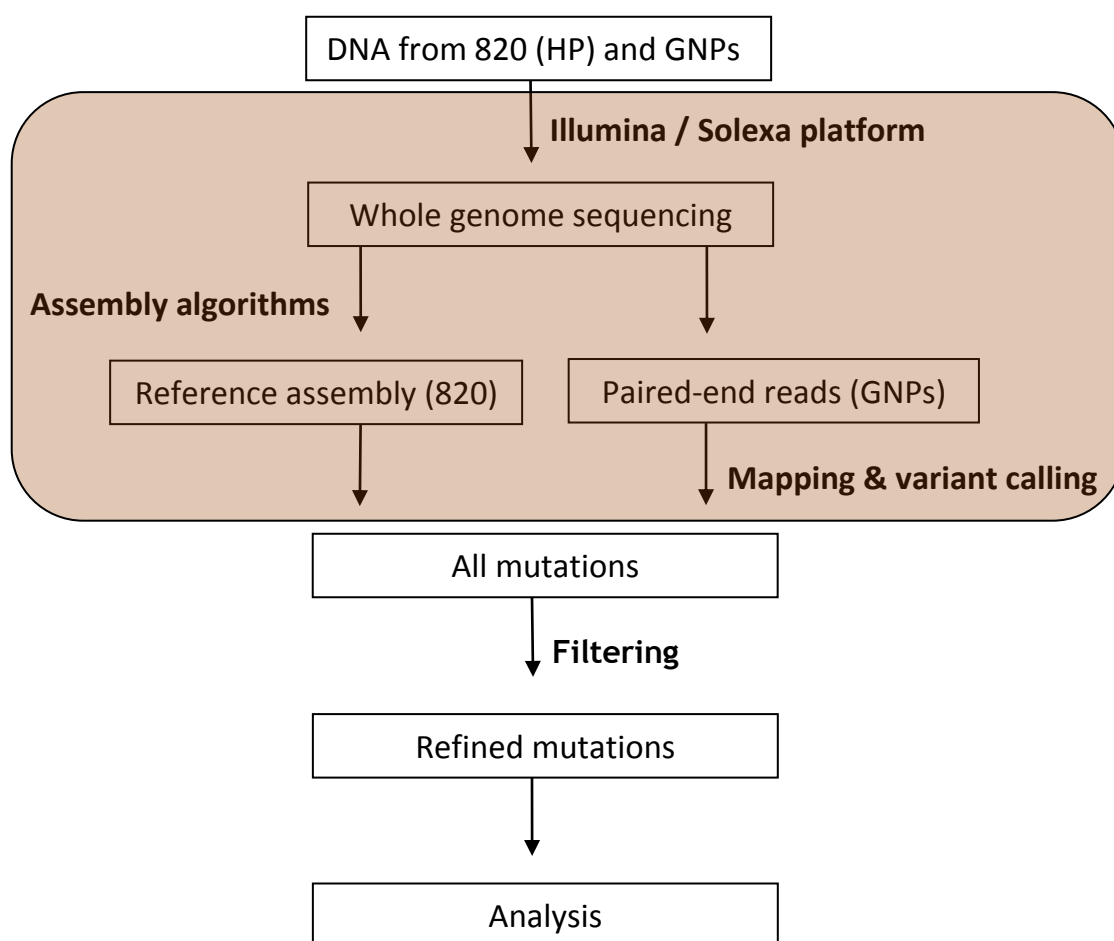
that had duplicated their DNA were apparent as a separate density in the FITC versus VioBlue density plots.

Any green fluorescent signal (which was supposed to be captured by the FL2 channel) that was leaking into the FL3 channel or vice versa was compensated in the appropriate channels by adjusting the amount of leaking signal intensity after acquisition of data.

Two negative controls, blood from an uninfected mouse and from a mouse infected with *P. berghei* HP ANKA line which is a known high producer of gametocytes (which were not fluorescently labelled), were used. These were used to precisely position the P1 gate in the FSC versus SSC plot to predominantly focus the red blood cell population. The samples were treated in the same way as the test samples and were also stained with Hoechst stain similarly.

## 2.6 Sequencing and bioinformatic methods

A flow diagram highlighting the major procedures performed from preparation of the DNA for sequencing up to the analysis of identified mutations is shown below (Figure 2-2). The work done by Wellcome Trust Sanger Institute (WTSI) has been shaded. Whole genome re-sequencing was performed by the pathogen sequencing unit at the WTSI. Assembly of the genomes, mapping of the reads onto the reference genomes and variant calling was performed by Dr Thomas Dan Otto of WTSI. This part of the work is mentioned only briefly here and only the work performed by me is detailed. A summary of basic algorithms used in generating a reference assembly, mapping and variant calling is presented in Table 2-1.



**Figure 2-2. A flow diagram representing the major pipeline for whole genome sequencing, assembly, variant calling and analysis of mutations.** The area shaded was performed at the WTSI.

### **2.6.1      *Preparation of DNA for whole genome sequencing***

Naïve phenylhydrazine-treated mice were infected with stabilates from WT parental line 820cl1m1cl1cl2 (820) and the three natural GNP lines, namely m9w21d (m9), m7w32d (m7), and m8w52d (m8). The mice were sacrificed once the parasitemia reached approximately 10% and blood from each of the infected mice was filtered with Plasmodipur filters (EuroProxima B.V., The Netherlands) and grown up to schizonts into gassed flasks at 37 °C overnight in a shaking incubator before being harvested for isolation of DNA. The DNA was quantified

on NanoDrop spectrophotometer (Fisher Scientific) and about 10 µg of the purified DNA from each of these lines shipped to WTSI under appropriate conditions for whole genome sequencing at WTSI.

### **2.6.2 *Illumina/Solexa whole genome sequencing and assembly of the GNP and their WT isogenic parental lines***

This section was entirely performed by Dr Thomas Dan Otto and his group at the WTSI. All the *de novo* GNP isogenic lines (lines m7, m8 and m9) along with their common parental line (820cl1m1cl1cl2) were sequenced by whole genome sequencing on the Illumina/Solexa sequencing platform using paired-end reads. In addition to this, the pre-existing GNP line K173cl1 and line 233 along with its gametocyte producer parent, line 234, were also sequenced similarly. A summary of basic algorithms used in generating a reference assembly, mapping and variant calling is presented in Table 2-1.

To sequence the clones 233, 820, m7, m8, and m9, libraries of 76 bp paired-end reads with an insert size range of 300-500 bp fragment length were generated following a PCR-free protocol (Quail et al., 2012). The libraries were sequenced using an Illumina Genome Analyser II with V4 chemistry.

A *de novo* assembly of reads was generated from the 820 parental clone using Velvet assembler (Zerbino and Birney, 2008) version 1.0.12 and the following parameters: -exp\_cov auto -min\_contig\_lgth 500 -cov\_cutoff 10 -ins\_length 350 -min\_pair\_count 20. Four hundred and seventeen supercontigs with an N50 of 240 kb were obtained. The assembly was processed as described in the post-assembly genome-improvement toolkit protocol (Swain et al., 2012). In short, scaffolds were ordered with ABACAS (Assefa et al., 2009) against the *P. berghei* ANKA reference genomes (GeneDB, version July 2010). This resulted in 16 pseudomolecules (14 chromosomes and 2 plastids) and a 'bin' of 100 contigs that could not be associated with a chromosome. Next, using scaffolds of at

least 1 kb as a substrate, IMAGE (Tsai et al., 2010) was used to close 469 (61 %) of the 774 sequencing gaps. Single base and small insertion or deletion (indel) errors were corrected using iCORN (Otto et al., 2010). This corrected 1067 single-base errors and 92 indels. Altogether, 1589 positions had heterozygous calls, which represented collapsed repeats, mostly in BIR genes.

Lastly, the annotation of the *P. berghei* ANKA reference genome was transferred onto the improved *P. berghei* 820 assembly using RATT (Otto et al., 2011) assembly option. In total, 4821 of the 4938 gene models were transferred correctly to the new genome, 35 were partially transferred and 82 gene models could not be transferred which were mostly in the sub-telomeric regions. Ninety seven percent (97%) of the bases in the *P. berghei* reference genome were finally covered by the completed 820cl1m1cl1cl2 assembly (available on: <ftp://ftp.sanger.ac.uk/pub/pathogens/Plasmodium/berghei/820/>).

### **2.6.3 Mapping and variant calling**

To call the variants, the software called SMALT (version 5.3; <http://www.sanger.ac.uk/resources/software/smalt/>) was used to map reads against the *P. berghei* ANKA reference and the generated 820 assembly. Variation was called with *pileup* and the *varfilter* (varFilter -D 2000; Quality >= 60) options from the SAMtools package (Li et al., 2009). For the reads mapped onto the 820 assembly, the variation of each clone, and concordance with other clones was analysed using a PERL script. For the reads mapped onto the ANKA reference genome, the script ignored variants that were called in all m7-m9 clones as well as 820, assuming them to be erroneous.

Table 2-1

Tools	What do they mean?	Purpose
<b>Velvet</b>	De novo sequence assembler	Creates the starting contigs from the reads
<b>ABACAS</b>	Algorithm Based Automatic Contiguation of Assembled Sequences	Orders contigs against a reference sequence
<b>IMAGE</b>	Iterative Mapping & Assembly for Gap Elimination	Identifies & closes gaps after contigs are ordered
<b>iCORN</b>	Iterative Correction Of Reference Nucleotides	Improves base-calling at specific positions, corrects actual sequence of the consensus genome that is being created
<b>RATT</b>	Rapid Annotation and Transfer Tool	Transfer of annotations from a reference onto a new consensus
<b>SMALT</b>	Sequence Mapping & Alignment Tool	Maps reads against a reference
<b>SAM Tools</b>	Sequence Alignment & Mapping	Calls sequence variants

Table 2-1. A summary of basic algorithms used in generating a reference assembly, mapping and variant calling

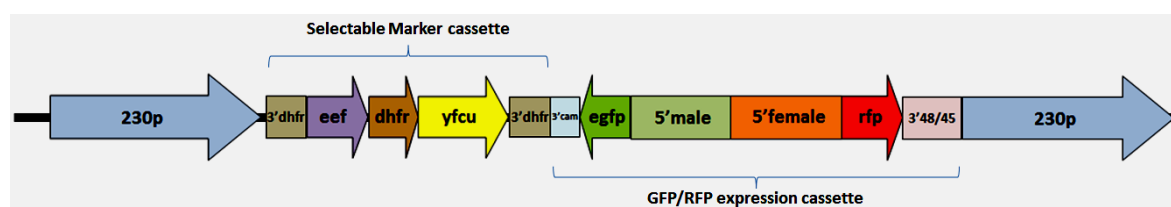
## 2.7 Parasite lines used in the study

### 2.7.1 *Lines used for assessing gametocytogenesis*

Cloned *P. berghei* parasite line PBANKA cl15cy1 820cl1m1cl1 (RMgm-164), a high gametocyte producer, was used as the parental line (Ponzi et al, 2009; Mair et al., 2010) for all genetic manipulations pertaining to understanding the



molecular basis of commitment to gametocytogenesis in malaria parasites. The line is transgenic for the production of RFP (in female gametocytes) and GFP (in male gametocytes) (Figure 2-3) and contains eGFP under the male gametocyte-specific promoter of putative dynein heavy chain gene (PBANKA\_041610) and 3'UTR of putative calmodulin gene (PBANKA\_101060) and RFP under the female gametocyte-specific promoter of an LCCL domain containing protein (PBANKA\_131950) and 3'UTR of the transmission blocking target antigen precursor P48/45 (PBANKA\_135960) at the p230p (PBANKA\_030600) replacement locus on chromosome 3. Thus, male gametocytes appear green and female gametocytes appear red under appropriate channels in flow cytometry and fluorescence microscopy. The line was re-cloned by the limiting dilution method to generate two pure parental line clones. One, the re-cloned line (PBANKA 820cl1m1cl1cl2) was used to infect a naïve Theiler's Original mouse as a common source parental line. At this stage the line was not checked for the presence of the selectable marker cassette as the line had been supplied on the understanding that negative selection had already been applied to this line.



**Figure 2-3. A cartoon of the parental WT line PBANKA 820cl1m1cl1 locus showing the drug selection marker (SM) cassette and GFP/RFP expression cassette positioned immediately adjacent at the 230p locus (light blue) on chromosome 3.** The SM cassette has *dhfr* (light brown) and *yfcu* (yellow) under control of pbeef1αa promoter (mentioned as *eef*; violet) and a pbdhfr 3'UTR (light grey). An extra copy of the pbdhfr 3'UTR is placed immediate upstream of the pbeef1αa promoter for SM recycling, if needed using negative selection. The GFP/RFP expression cassette has *egfp* (bright green) under the male gametocyte-specific promoter (putative dynein heavy chain gene; PBANKA\_041610; light green) and the 3'UTR of putative calmodulin gene (PBANKA\_101060; light blue; mentioned as 3'cam) together with *rfp* (red) under the female gametocyte-specific promoter (LCCL domain containing protein; PBANKA\_131950; orange) and the 3'UTR of the transmission blocking target antigen precursor P48/45 (PBANKA\_135960; light pink).

## 2.7.2 Lines used as negative control

The high gametocyte producer (HP) line of *P. berghei* ANKA was used as a negative control for fluorescence during FACS analysis and other experiments. In this line, the parasites stages are not labelled with any reporter proteins. Thus,

any positive fluorescent signal detected on FACS using *P. berghei* HP ANKA represented a background noise and hence was discounted from the relevant readings.

### **2.7.3      *Lines used for reference sequencing***

DNA from a number of related high gametocyte producer and low or pre-existing gametocyte non-producer *P. berghei* lines were used for either whole genome re-sequencing or direct Sanger sequencing to detect any lineage-specific mutations. Mutations associated with the GNP phenotype in the isogenic successors of the parental 820 line were compared with mutations (if any) discovered in these reference lines. These lines included the line 234 (high gametocyte producer), and lines 233 and K173 (pre-existing GNP lines).

### **3    Generation of *de novo* gametocyte non-producer mutants**

### 3.1 Introduction

In *Plasmodium berghei*, during each of the asexual multiplication cycles, roughly between 10% and 20% of the asexually dividing parasites differentiate to the sexual precursors - the gametocytes (Janse and Waters, 1995). It is also well documented that *Plasmodium* parasites, otherwise producing an optimum number of gametocytes during their asexual replicative cycle in mammalian hosts, are known to spontaneously and irreversibly lose the ability to produce gametocytes under conditions of repetitive mechanical passages between their natural mammalian hosts in the absence of a vector (Micks, 1947; Janse et al., 1989; Dearsly et al., 1990; Janse et al., 1992). In *P. berghei*, these gametocyte non-producer (GNP) mutants (or low gametocyte producer variants) could spontaneously arise from their high gametocyte producer parent line as a result of either unintentional prolonged continuous maintenance of these gametocyte producer parasite lines in mice without being exposed to mosquitoes, for example clone 1 of the K173 lineage (Vincke and Lips, 1948; Janse et al., 1989; Janse et al., 1992) and clone 233L derived from the parental clone 234L of the ANKA lineage (Killick-Kendrick, 1974; Dearsly et al., 1990); or as a planned and intended experimental outcome generating *de novo* GNP parasite lines (Micks, 1947; Janse et al., 1992). Gametocyteless clones have also been generated in *P. falciparum* including clone C10 from isolate 1776 (Shirley et al., 1990; Day et al., 1993) and clones F12 and E4 from isolate 3D7 (Alano et al., 1995).

Isogenic *de novo* GNP mutant lines are critical for forward genetics studies aimed at identifying the molecular basis of the resultant GNP phenotype. The underlying reason is straightforward - because the pre-existing GNP mutant lines remain an unknown number of generations apart from their gametocyte producing parents, and hence have the likelihood of acquiring and accumulating mutations over time. Any comparative mutational analysis between pre-existing GNP lines and their high gametocyte producer parent lines is expected to have a much higher noise to signal ratio in contrast to a similar analysis involving the *de novo* GNP lines and its high gametocyte producer parent lines. Although the rate of accumulation of mutations over time is low in *Plasmodium*, ( $1.7 \times 10^{-9}$  and  $4.6 \times 10^{-9}$  mutations per base pair per generation in the absence of drug pressure and for Atovaquone (ATQ)-resistant clones, respectively in *P. falciparum* 3D7),

any single accumulated mutation would tend to confound the findings of the actual result (Bopp et al., 2013). Thus, in the present study, attempts were made to generate 10 independent isogenic *de novo* GNP mutant lines from a single cloned high gametocyte producer parent line to provide enough statistical power so that any differences in the mutational profiles between the parent and GNP lines could be truly attributed to the differences in their phenotypic profile.

## 3.2 Flow-cytometry (FACS) to monitor gametocytemia

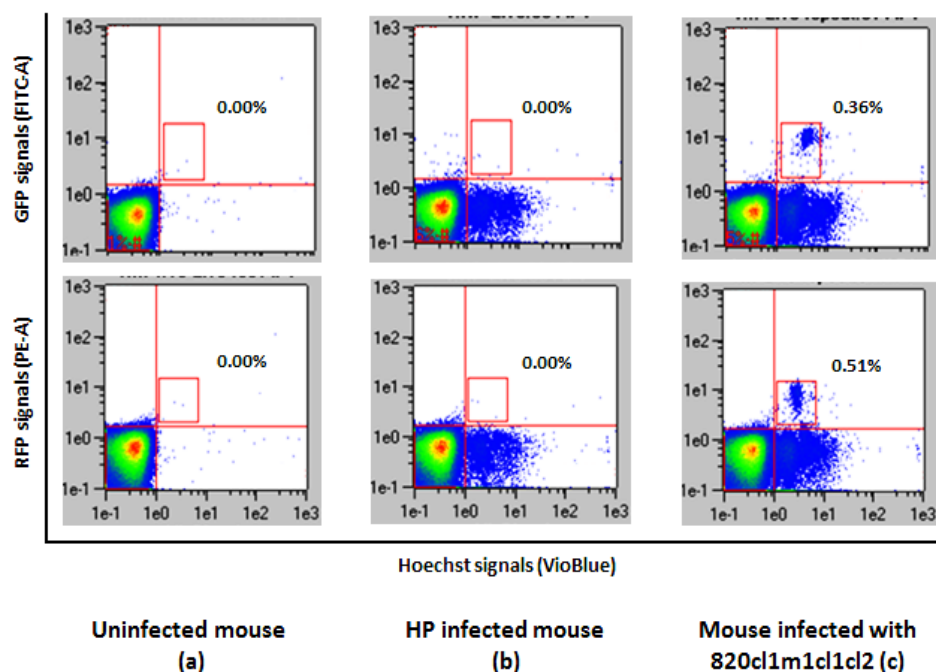
Weekly monitoring of total parasitemia and gametocytemia was done by flow-cytometry (FACS) using 1-2 drops of peripheral blood from each of the ten isogenic passage lines. Gametocytemia was demonstrated in the presence of appropriate controls (Figure 3-1). A negative control for infection (Hoechst-stained uninfected mouse blood), negative control for green and red fluorescent signals in presence of infection (Hoechst-stained blood from a non-fluorescent high gametocyte producer line infected mouse) and a positive control for green and red fluorescent signals in the presence of infection (Hoechst-stained blood from parental 820cl1m1cl1cl2 line) were included in all the FACS analyses.

## 3.3 Asexual maintenance of ten isogenic lines

Briefly, ten independent but parallel isogenic lines (all derived from a single cloned high gametocyte producer line 820cl1m1cl1cl2) were maintained asexually by repeated mechanical passage between mice in the absence of mosquito bites for ~52 weeks or the development of a gametocyte non-producer phenotype, whichever was earlier (See Material and Methods; section 2.1.1.3). Based on earlier experiments, (Janse et al, 1992), it was reasonable to expect that eventually by the end of the one year passage period, most, if not all, lines would have lost their capacity to produce gametocytes.

However, in the current study, out of the 10 potential lines, only 3 stopped producing gametocytes (lines 9, 7 and 8 - in that order) at the end of

the study period. The remaining seven were still producing a sub-optimal number of gametocytes (although at varied and reduced levels).



**Figure 3-1. Flow-cytometry (FACS) plots showing gametocytemia in the experimental controls.** On the horizontal axis are shown the cells positive for Hoechst-33258 signals as detected in the VioBlue channel signifying the parasite infected RBCs. The vertical axis shows the GFP positive cells (male gametocytes; upper panel) and the RFP positive cells (female gametocytes; lower panel) as detected in the FITC-A and PE-A channels, respectively. No Hoechst-positive and GFP/RFP positive cells were detected for blood drawn from uninfected mouse (a) whereas significant Hoechst positive cells were observed for the HP infected mouse with no GFP/RFP positive cells (b) demonstrating that negative controls for Hoechst and GFP/RFP. The positive control (blood from mouse infected with WT parental line 820cl1m1cl1cl2; c) shows both the Hoechst positive cells (signifying the presence of infection) and the GFP/RFP positive cells (signifying the presence of male and female gametocytes at 0.36% and 0.51% of the total parasitemia, respectively).

Although the experiment was planned for 52 weeks from the start of the first passage, further passages of the low gametocyte producer lines could be attempted in future by reactivating the weekly cryopreserved stabilates of the relevant lines.

### 3.4 Growth kinetics among ten isogenic lines

For each of the 10 parasite lines, growth of the parasites for each week in terms of total parasitemia (percentage of infected RBCs out of total RBCs) and gametocytemia (percentage of gametocytes out of total infected RBCs) as

detected by FACS was plotted along the 52-weeks period of asexual multiplication (passages). The figures (numbered 3-2 to 3-11) below show the kinetics of growth in each of the ten lines. The total parasitemia is represented on the secondary Y-axis by the bold dark blue smooth line and the gametocytemia is represented along the primary Y-axis by bold grey smooth line. The gametocytemia was further subdivided into male gametocytemia (light blue smooth line) and female gametocytemia (pink smooth line) and is represented on the primary Y-axis. The X-axis shows the period of asexual multiplication (passage) in weeks.

Each of the smooth lines is the best-fit polynomial regression line (order 6) to the corresponding individual weekly observations joined by straight lines. In order to be able to detect very low number of gametocytes (even between 0% and 1% of the total parasitemia), the total parasitemia of all passaged lines was allowed to rise anywhere between 5% and 10%. However, during the initial phase of the experiment (initial 4-8 weeks), due to optimization of the protocol, total parasitemia for some of the lines could not be strictly maintained between the desired level of 5%-10% at passage and was a little lower than 5% thus resulting in some difficulty in interpreting the gametocytemia during that period of study. None of the 10 parasite lines lost the ability to produce gametocytes during this initial experimental phase. But beyond that period, the total parasitemia was stably maintained as desired. The weekly data for total parasitemia and gametocytemia are shown in Appendices (Tables A1-A10).

### 3.4.1 Kinetics of growth in line 1

On average, the total parasitemia in line 1 was maintained between 5% and 10% barring week number 9 to 18, where it dipped to just below 4%. The gametocytemia showed the initial fluctuating phase (from week 1 to week 24) where it ranged between 10% and 20% at passage. This was followed by a declining phase where the gametocytemia declined to ~3% (from week 25 to week 34). The third phase in gametocytemia kinetics (week 35 to week 52) showed a fluctuation between 1% and 3% (which is strikingly at a lower value and within a narrower range as compared to the initial fluctuating phase in the gametocytemia kinetics in this line). The experiment was terminated after the parasites were passaged for 52 weeks as this was the planned end point of the experiment, should there be no time point where the gametocytemia was not detected (Figure 3-2).

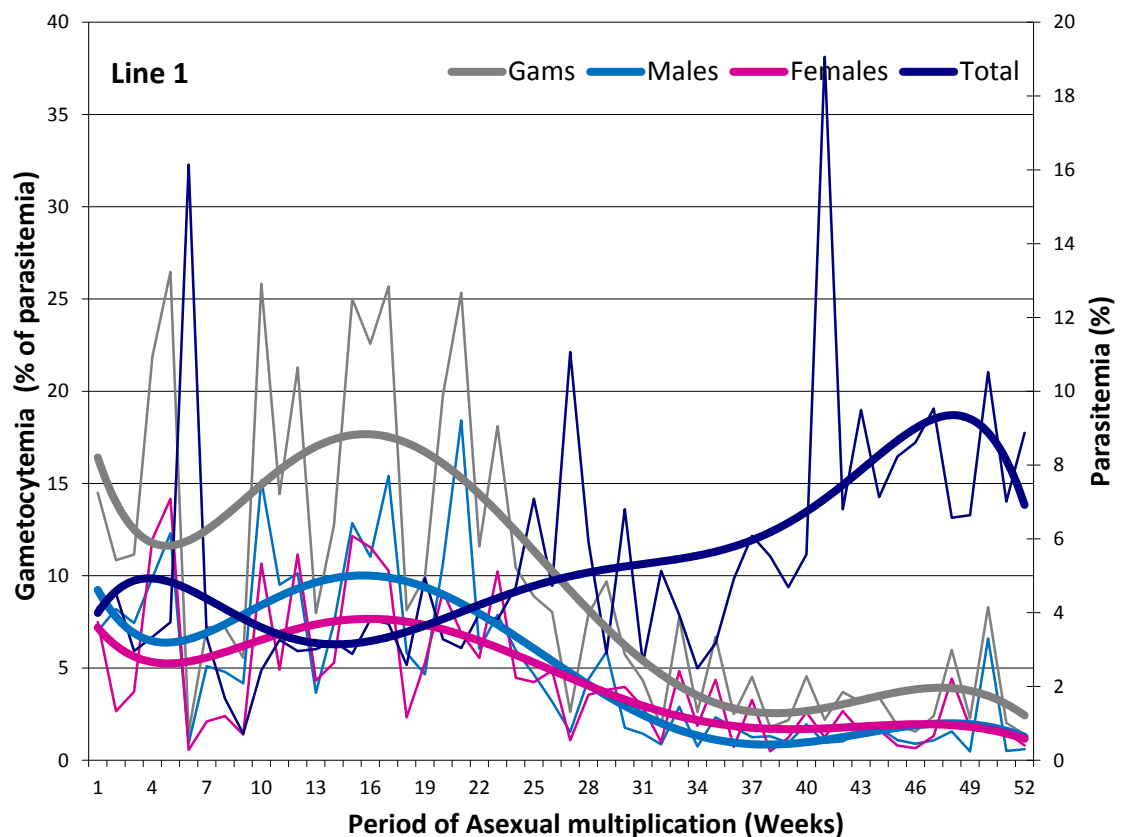


Figure 3-2. Parasitemia of the serially passaged line 1 at passage.



### 3.4.2 Kinetics of growth in line 2

Line 2 was passaged for a total of 42 weeks only as there were 4 occasions during the experiment when the passage was not successful as no parasites could be detected in the passaged naïve mice. Also, except for the initial 8 weeks in the passage period, the total parasitemia was maintained at the planned level ranging between 5% and 10% at passage. The gametocytemia followed a similar 3-phased trend as observed in line 1 but this was less prominent and within a narrower range than line 1. The first phase (high fluctuating gametocytemia; between 10% and 15%) lasted for the initial 17 weeks followed by a declining phase (week 18 to week 29) and finally a low fluctuating phase (week 30 to week 47) when the gametocytemia ranged between 2% and 3% (Figure 3-3).

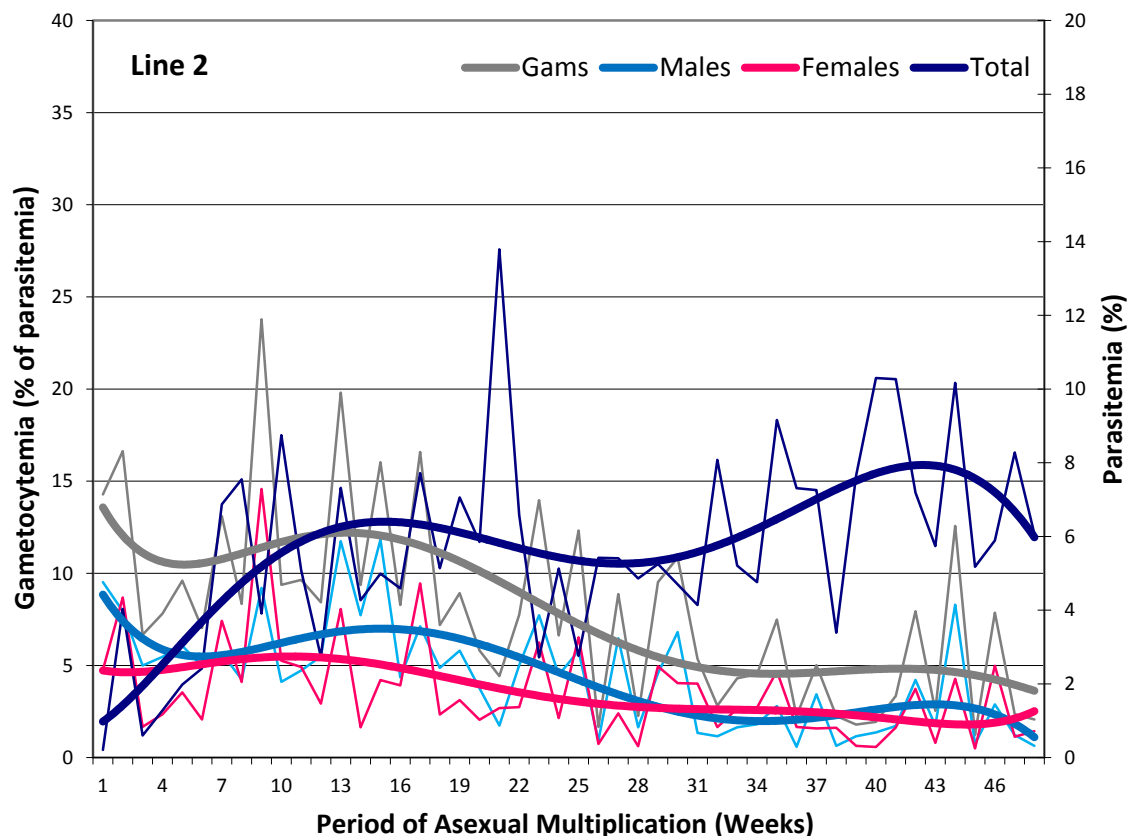


Figure 3-3. Parasitemia of the serially passaged line 2 at passage.

### 3.4.3 Kinetics of growth in line 3

Line 3 was maintained in asexual multiplication in mice for up to 53 weeks. It was only once during the total passage period of 53 weeks that it was not possible to mechanically passage the parasites. The total parasitemia was between 3% and 5% between weeks 3 and 23 otherwise the parasitemia at passage was maintained between 5% and 10%. The fluctuations in gametocytemia were clearly evident as three visually distinct phases and hence the gametocytemia kinetics observed in this line was used as a prototype to explain the trend of gametocytemia over the total period of asexually multiplying parasites in the rest of the seven lines that were still gametocytes producers at the end of the experiment. For the initial 25 weeks, the gametocytemia fluctuated between 10% and 15% which was followed by a steady decline to ~1% over the next 14 weeks. From week 39 to week 53, the gametocytemia fluctuated at a lower (than initial) level and ranged between 1% and 2%, thus demarcating the lower fluctuating phase (Figure 3-4).

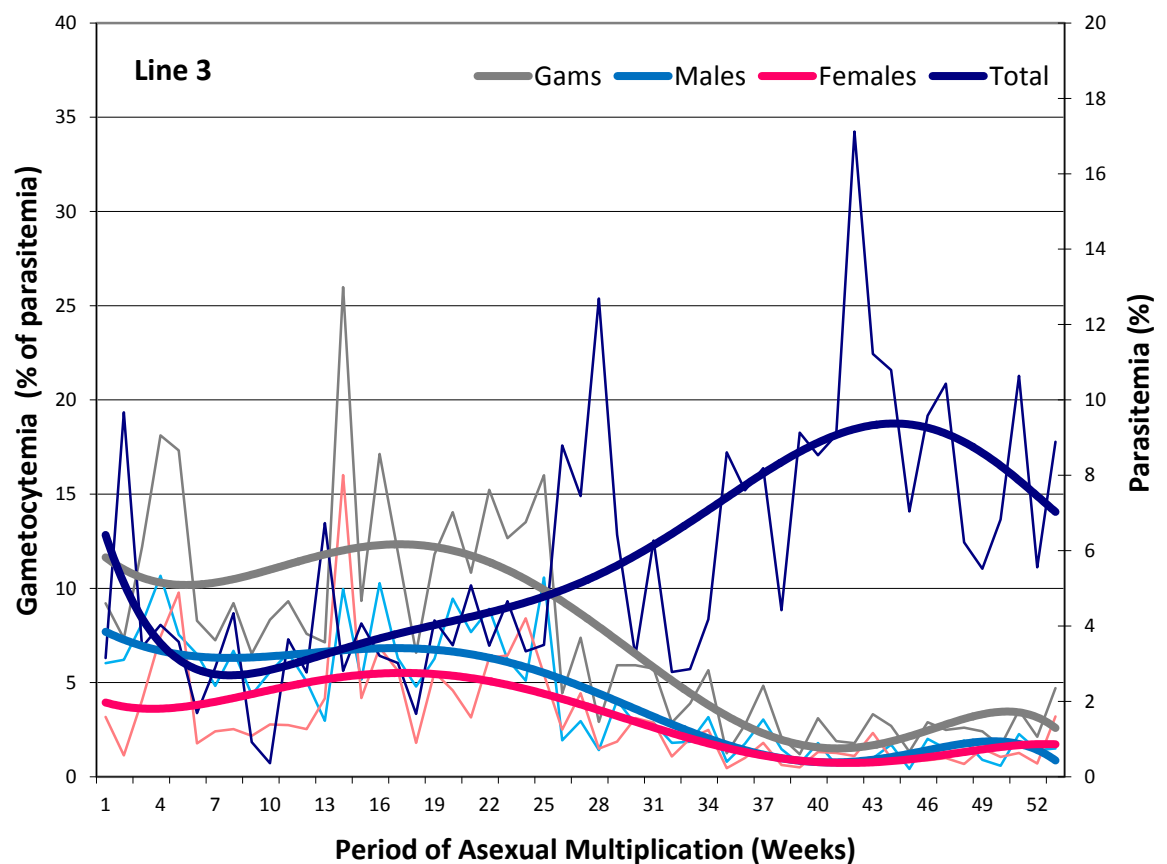


Figure 3-4. Parasitemia of the serially passaged line 3 at passage.

### 3.4.4 Kinetics of growth in line 4

This line was maintained in asexual multiplication for a total of 51 weeks only as it was necessary to reactivate the line 6 times during the course of the experiment. It is for this reason that the total parasitemia could not be maintained between the desired level of 5% and 10% at passage for the first 33 weeks. The gametocytemia showed a clear high fluctuating phase (between 9% and 16%) spanning the initial 31 weeks, followed by a declining phase where the gametocytemia declined to ~2% by week 38. The last phase was marked by a low(er) fluctuating gametocytemia for the remaining 13 weeks where it ranged from less than 1% to 2% (Figure 3-5).

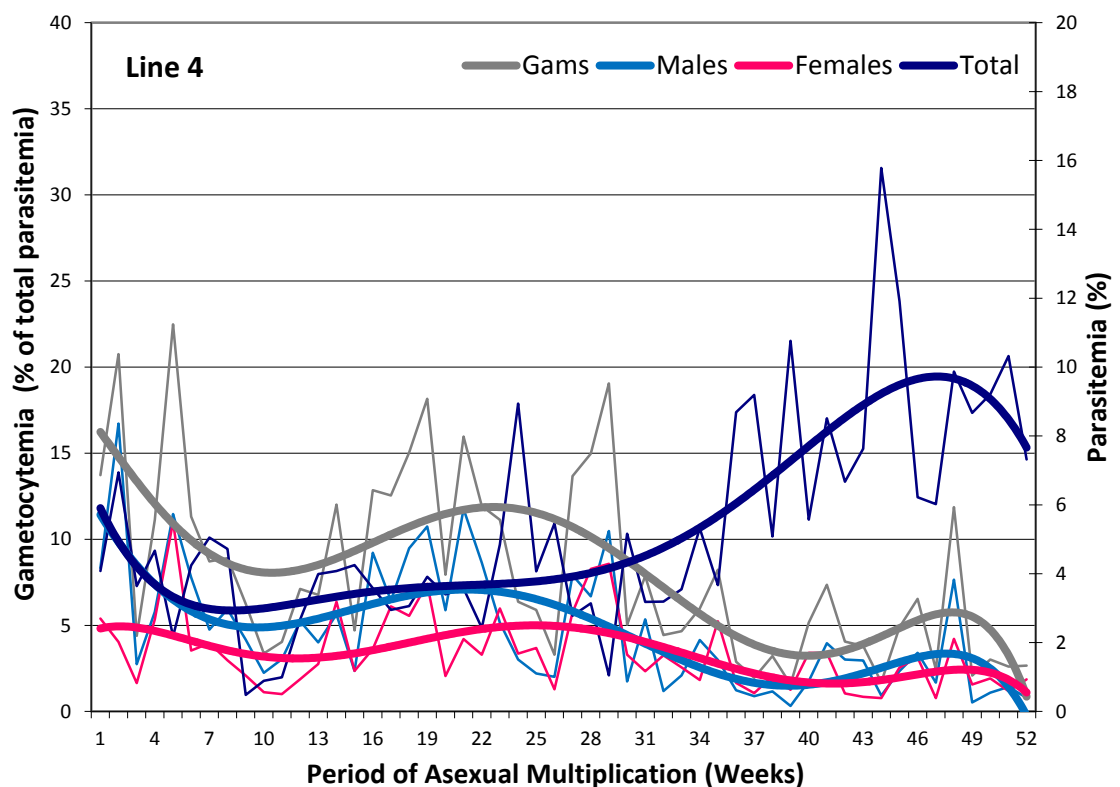


Figure 3-5. Parasitemia of the serially passaged line 4 at passage.

### 3.4.5 Kinetics of growth in line 5

Line 5 could be passaged asexually for up to week 51 with but was reactivated from stablate 7 times during this period. Barring the initial 15 weeks, the average total parasitemia was maintained within the desired range of 5% to 10% at passage. The fluctuations in the early fluctuating gametocytemia phase (initial 25 weeks) were quite prominent (ranging from ~7% to ~15%). The declining phase (week 26 to week 31) was followed by a rather stable low gametocytemia phase (week 32 to week 51) where the average fluctuations in gametocytemia could be barely appreciated. This is in sharp contrast to other low gametocyte producing lines generated in this study where this phase had fluctuating gametocytemia (but at a lower level than the early fluctuating phase) (Figure 3-6).

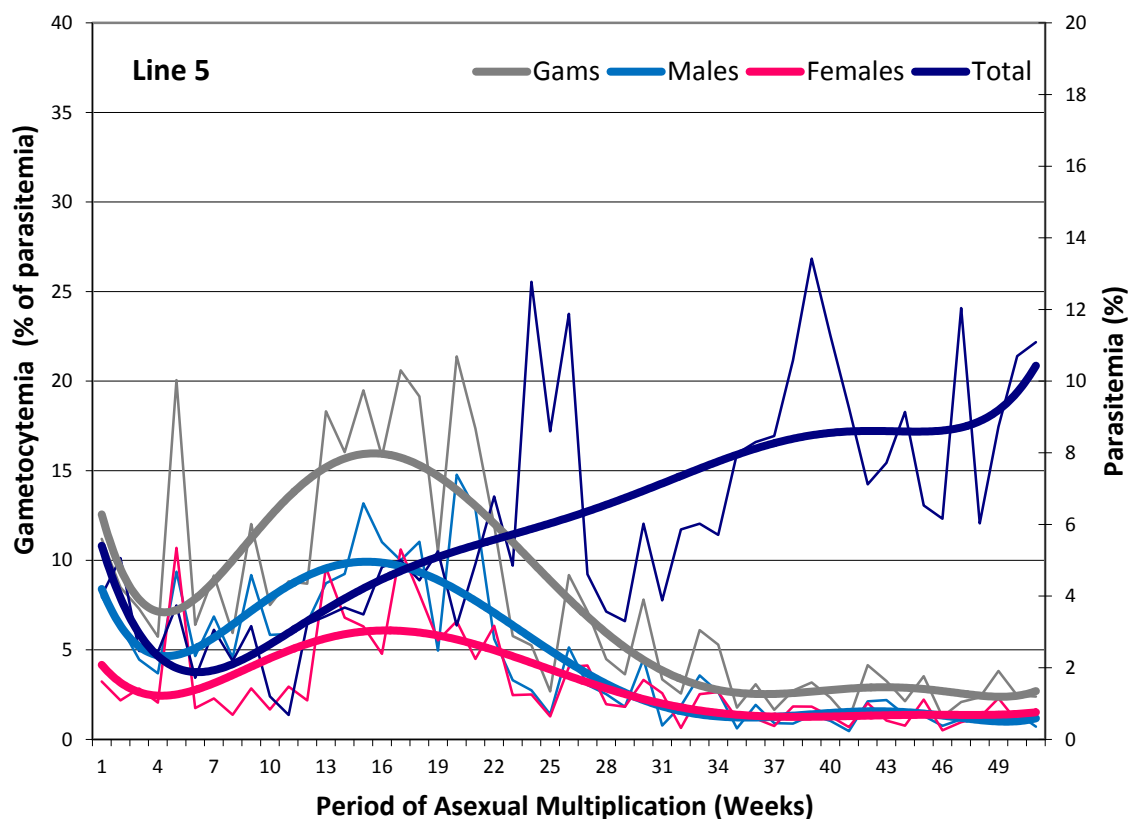


Figure 3-6. Parasitemia of the serially passaged line 5 at passage.

### 3.4.6 Kinetics of growth in line 6

This line could be passaged for the desired period of 53 weeks in total but with 3 reactivations from stabilates during the period. Similar to the other lines, the total parasitemia could not be maintained in the desired range of 5% to 10% at passage during the initial 18 weeks. The early fluctuating gametocytemia phase (up to initial 30 weeks) was unremarkable and so was the declining phase (week 31 to week 39). However, the late fluctuating gametocytemia phase (week 40 to week 53) was similar to that observed in line 5 in terms of the range of fluctuations in gametocytemia (Figure 3-7).

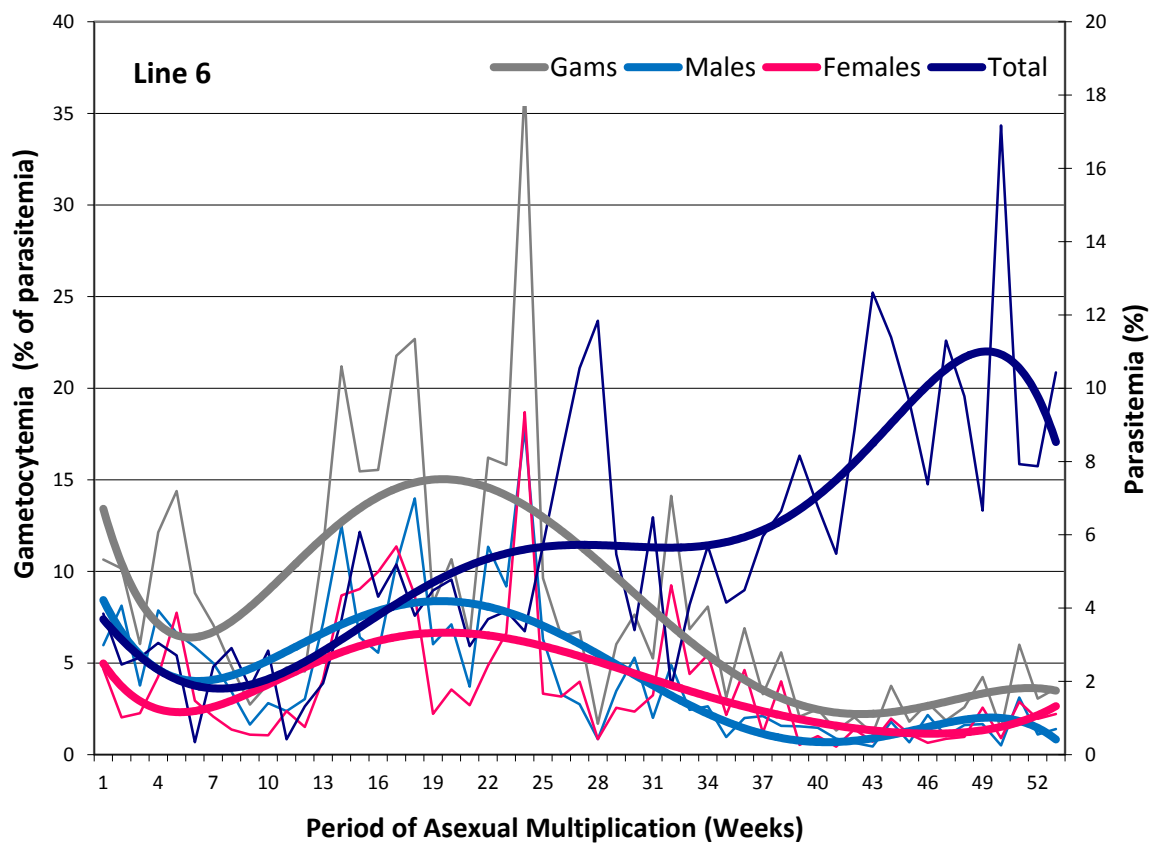
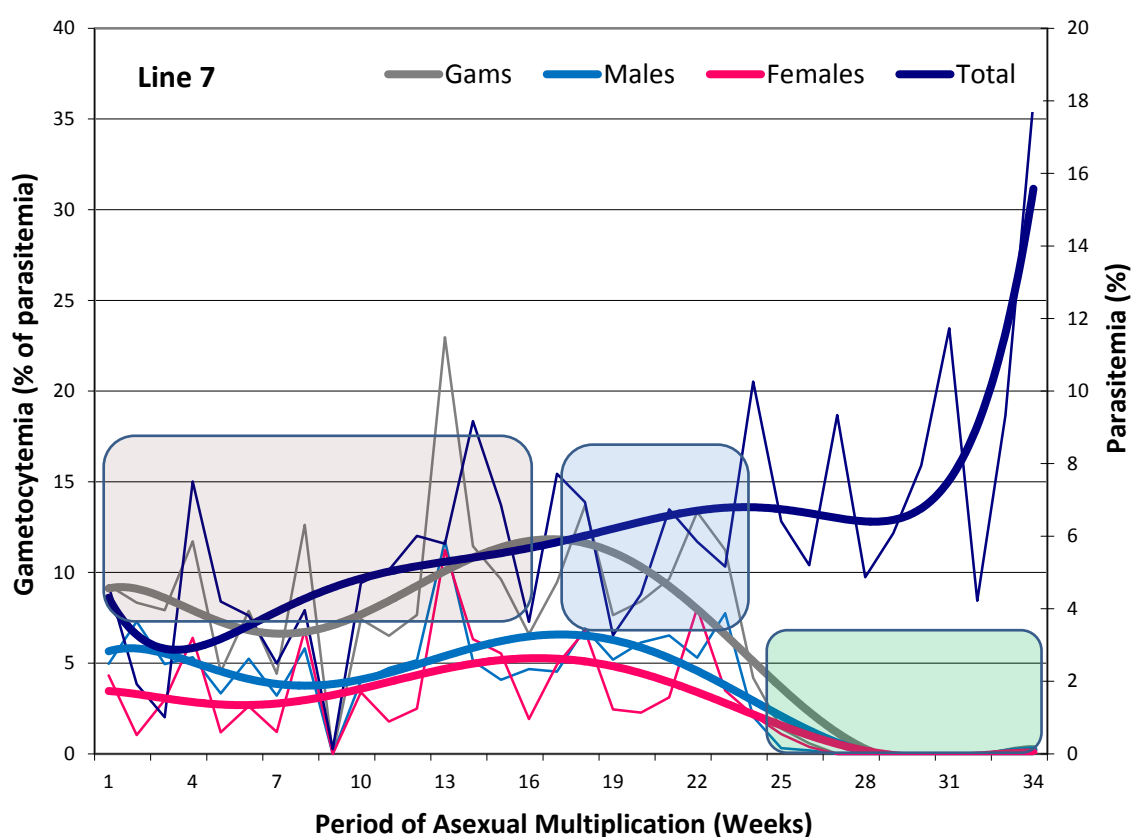


Figure 3-7. Parasitemia of the serially passaged line 6 at passage.

### 3.4.7 Kinetics of growth in line 7

Line 7, along with lines 8 and 9 (see below) ended up being a gametocyte non-producer (GNP). Line 7 became GNP after 32 weeks of continuous asexual multiplication with 5 reactivations in between. This line was maintained in

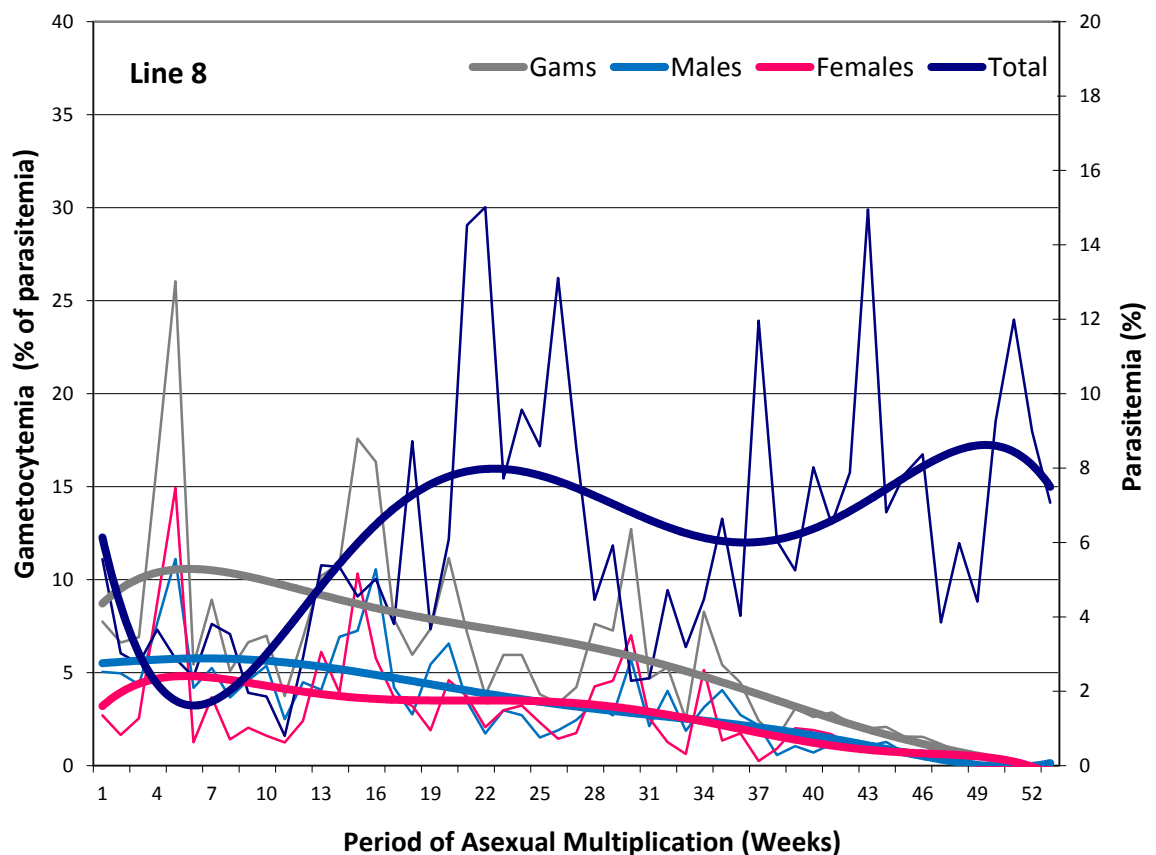
asexual multiplication only for 33 weeks as it first demonstrated absence of gametocytes at week 28 and hence was only continued for further 5 weeks in order to confirm the phenotype. As observed with the other lines, the total parasitemia was below the desired lower level of 5% at passage for the initial 10 weeks. The total parasitemia was deliberately raised to beyond 10% during the last 2 weeks of passage to allow the detection, if any, of even the minimal proportion of gametocytes that could have been present. The early fluctuating gametocytemia phase (first 17 weeks) was unremarkable. However, the declining gametocytemia phase (week 18 to week 27) was remarkable in sharp contrast to the other lines which still showed gametocytes (although at a sub-optimal level) at the end of the period of asexual passages - instead of terminating as a low fluctuating gametocytemia phase, the decline in gametocytes continued until a point (week 28) where no gametocytes could be detected. This marked the first evidence of a loss of gametocytogenesis and hence was monitored for a further period of 5 weeks. At the end of 32 weekly passages, this line was finally categorised as a GNP line (Figure 3-8).



**Figure 3-8. Parasitemia of the serially passaged line 7 at passage.** The coloured transparent rectangles (pink, blue and green) indicate different phases of the gametocytemia kinetics in the line. The tails (pink and blue) observed for gametocytemia after passage week 31 is due to the smoothed polynomial being fitted for the kinetics and do not represent actual gametocytemia.

### 3.4.8 Kinetics of growth in line 8

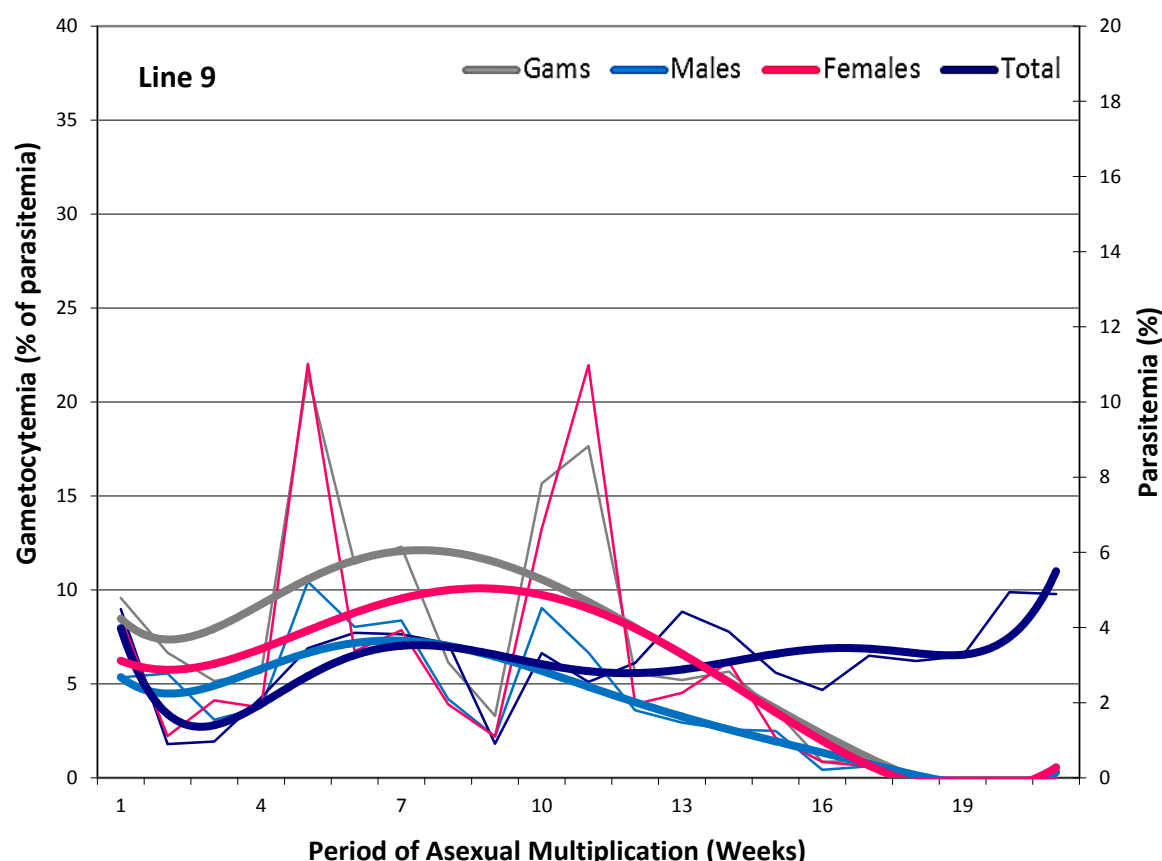
Other than the initial 14 weeks of passage, line 8 was maintained in asexual reproduction with a total parasitemia between 5% and 10% at passage. The line was passaged for 53 weeks with 5 intermediate reactivations from stabilates. In terms of gametocytemia kinetics, this line is unique to the other lines included in this experiment in terms of (a) it ultimately ended as being a gametocyte non-producer line and, more intriguingly, (b) was inconspicuous for the distinct gametocytemia phases during the passage period. The polynomial trend of gametocytemia declined steadily from week 7 of the weekly passages and continued to decline to no gametocytemia level, without being interrupted by low gametocytemia fluctuations (as observed in the other 9 lines) (Figure 2-30). This atypical trend observed could have implications for understanding the way the whole process of gametocytogenesis is controlled and regulated (see proposed models later) (Figure 3-9).



**Figure 3-9. Parasitemia of the serially passaged line 8 at passage.** The tails (pink and blue) observed for gametocytemia after passage week 52 is due to the smoothed polynomial being fitted for the kinetics and do not represent actual gametocytemia.

### 3.4.9 Kinetics of growth in line 9

Line 9 was the first line to lose the capacity to form gametocytes happening as early as week 18 of the continuous weekly passages. The total parasitemia in this line was never maintained within the desired range and was ~4% at passage for the majority of the experiment. The gametocytemia kinetics were not much different to that observed in line 8 (other than the initial fluctuations up to week 10) in terms of both the final end result as well as the absence of the late fluctuations in the gametocytemia levels. The first evidence of loss of gametocytogenesis was observed at week 18 and hence the line was only followed for a further 4 weeks. Line 9 mouse-to-mouse passage was interrupted 5 times during its passage period where it involved reactivation of the line with the help of the cryopreserved stabilates of the most recent passage (Figure 3-10).



**Figure 3-10. Parasitemia of the serially passaged line 9 at passage.** The tails (pink and blue) observed for gametocytemia after passage week 20 is due to the smoothed polynomial being fitted for the kinetics and do not represent actual gametocytemia.



### 3.4.10 Kinetics of growth in line 10

Line 10 was passaged for a total of 53 weeks but was recovered 3 times from the stabilates during this passage period. The total parasitemia was maintained within the desired range - between 5% and 10% at passage for most of the passage period except the initial 13 weeks, as in most of the lines. The gametocytemia followed the typical trend of three visually distinct phases - optimal fluctuating gametocytes (between 5% and 10%; up to week 26), declining gametocytes (week 27 to 35) and sub-optimal fluctuating gametocytes (between 1% and 3%; week 36 to 53) (Figure 3-11).

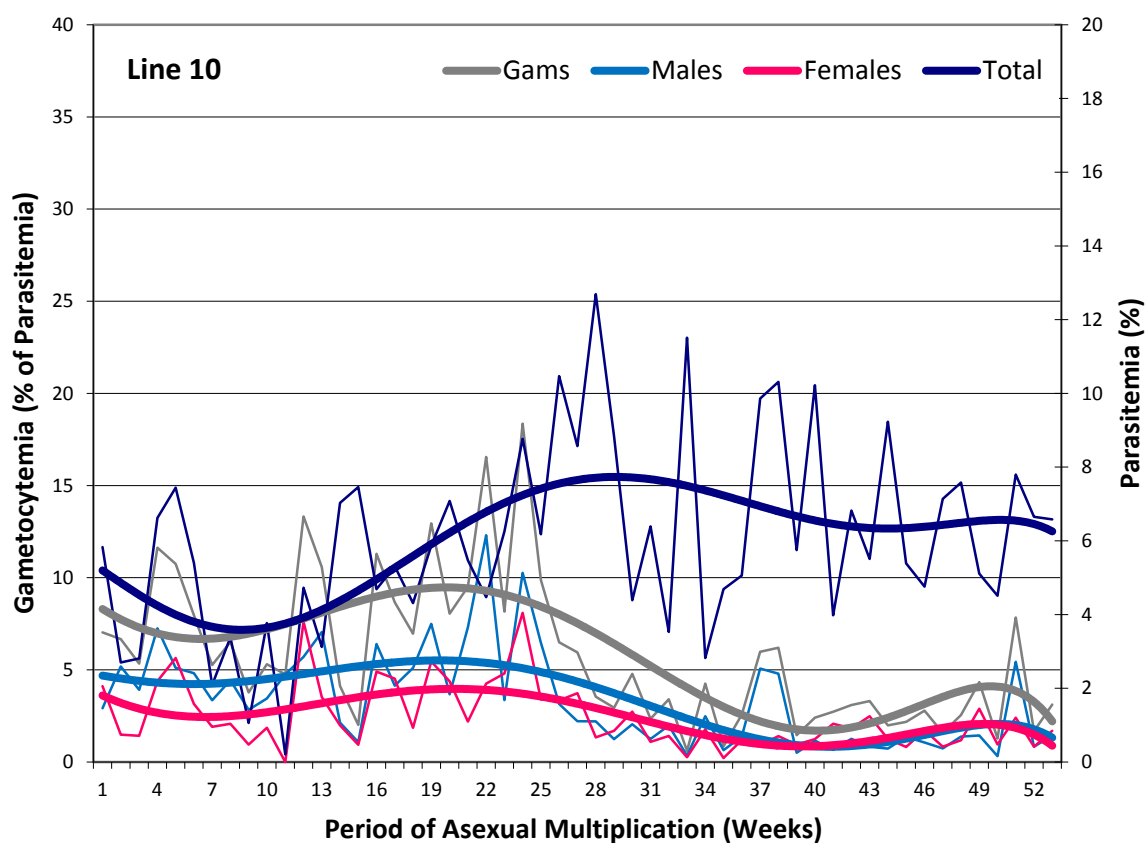
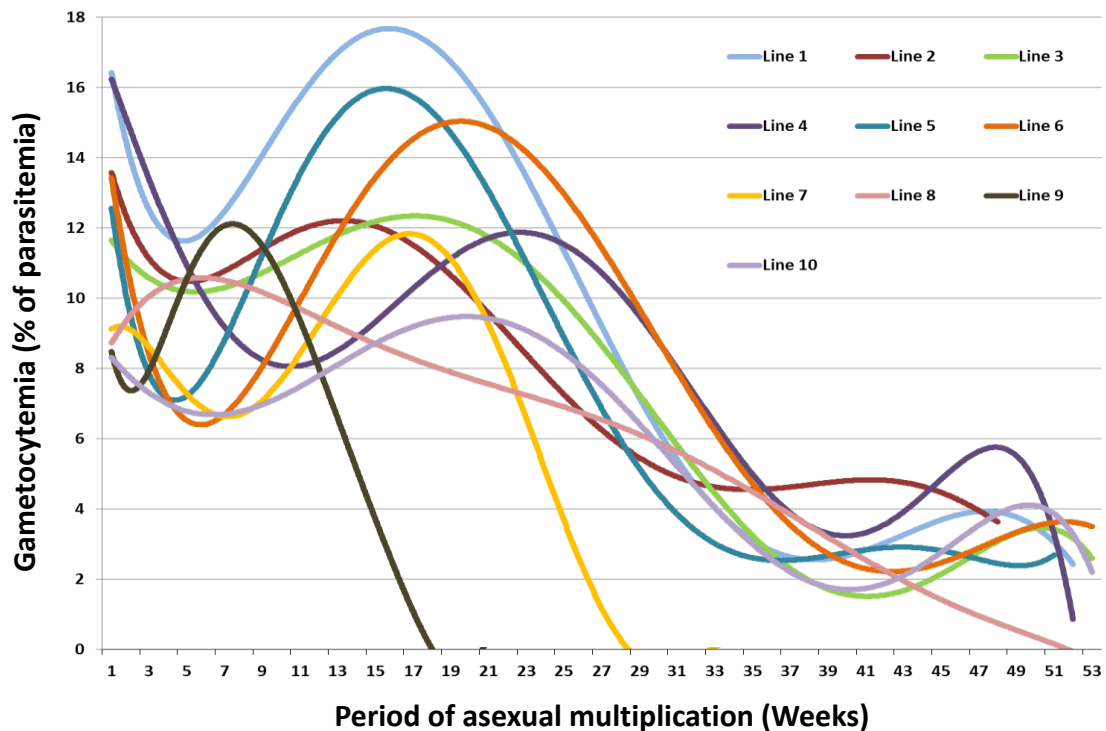


Figure 3-11. Parasitemia of the serially passaged line 10 at passage.

### 3.5 *Lines 9, 7 and 8 did not produce gametocytes*

As evident from the parasitemias and gametocytemia of the serially passaged lines in figures 3-3 to 3-12 above, the gametocytemia declined in all the ten lines studied but reached to zero in only three lines (Figure 3-12).



**Figure 3-12. Gametocytemia of the ten serially passaged lines at passage.** Each of the coloured smooth lines is the best-fit polynomial regression line (order 6) to the corresponding individual weekly observations (not shown here due to clarity of the figure). As evident, the gametocytemia declined in all the ten lines studied but reached to zero only in lines 9, 7 and 8 in that chronological order.

Thus, lines 9, 7 and 8 evolved a gametocyte non-producer phenotype (in that chronological order) on week 21, 32 and 52, respectively. The absence of gametocytes in these GNP lines was ascertained by:

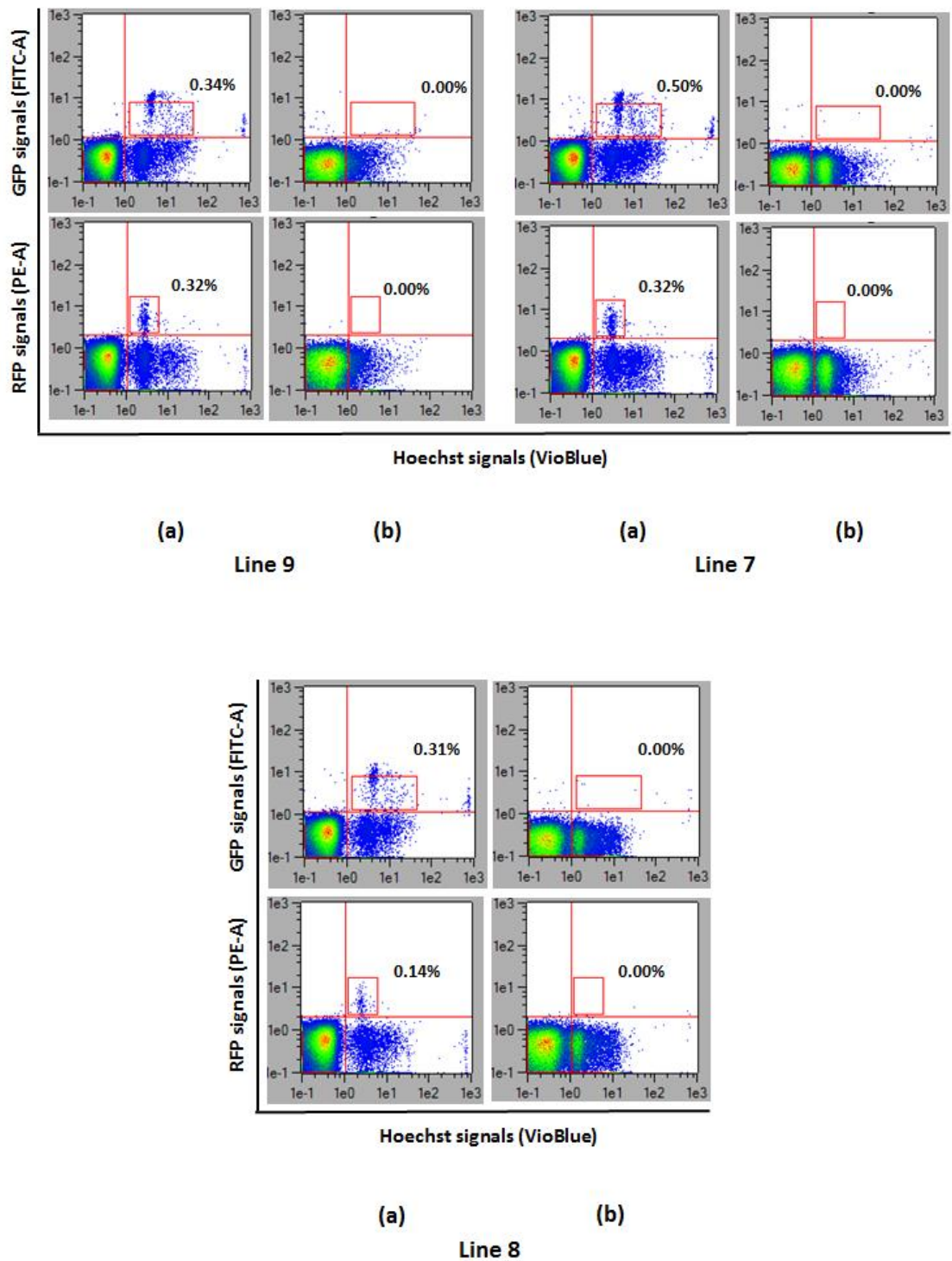
#### 3.5.1 Giemsa-stained peripheral blood smears

No gametocytes were detected on Giemsa-stained smears (out of the 1500 parasites examined per line) from lines 7, 8 and 9 throughout the 4-week period starting from the point where absence of gametocytes was first detected by

FACS. This was in contrast to 142 gametocytes observed in the WT parental line 820 (data not shown).

### 3.5.2 Flow cytometry analyses

Absence of gametocytogenesis was also established by flow cytometry (FACS) based on the absence of red and green fluorescent signals for female and male gametocytes, respectively for at least four consecutive weekly passages after the absence of fluorescent signals was first detected by FACS. The absence of gametocytes was demonstrated in the presence of appropriate controls (Figure 3-1). The first line to lose gametocytogenesis was line 9, stopping to produce gametocytes at week 21 of the repeated passages. The next in the chronology was line 7 which lost gametocytogenesis at week 32 and the last line to become a GNP was line 8 which was conclusively shown by FACS to stop gametocytogenesis at week 52 of the continuous asexual growth. The initial and final gametocytemia of the three GNP lines is shown in Figure 3-13. The FACS observations together with that from Giemsa-stained peripheral smears confirm the absence of any detectable gametocytes in all the three *de novo* GNP lines generated in this study.



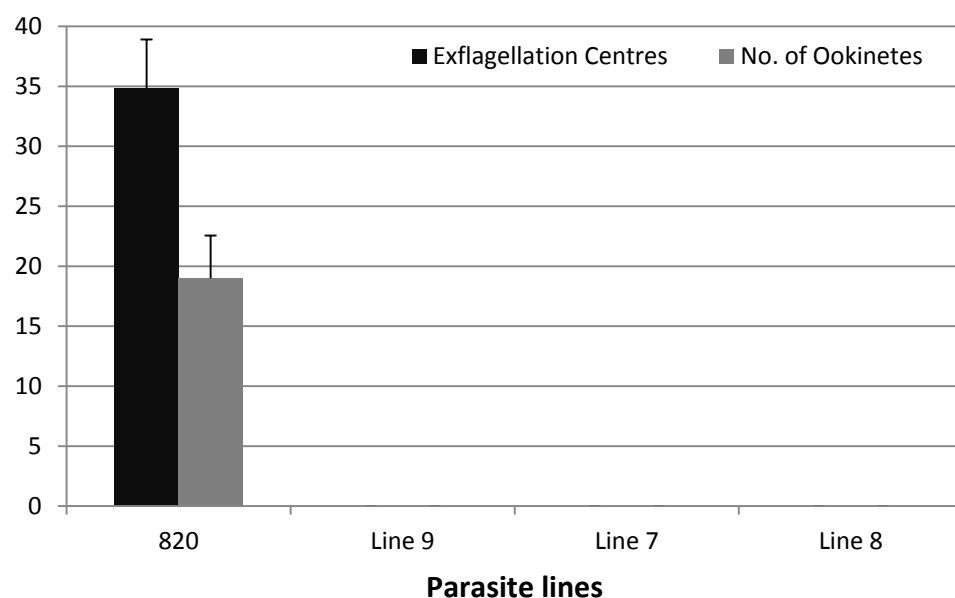
**Figure 3-13. FACS analysis of Hoechst-stained blood from each of the *de novo* GNP lines as they arose – line 9, line 7 and line 8.** X-axis: Hoechst signals as detected in the VioBlue channel; Y-axis: GFP signals as detected in the FITC-A channel (upper) & RFP signals as detected in the PE-A channel (lower plots). The gametocytes are gated (red rectangle) in each plot. A comparative male and female gametocytemia (%), GFP and RFP signals, respectively, is shown between the values at the start of the experiment/passage (week 1; a) and at the confirmation of their GNP status (weeks 21, 34 and 53 for line 9, line 7 and line 8, respectively; b). As clearly seen, these FACS findings confirm the absence of any detectable gametocytemia in each of the GNP lines (as compared to the controls depicted in Figure 3-1).

### 3.5.3 Exflagellation assays

The evidence for loss of gametocytogenesis was further supported by the exflagellation assays which detect the presence of exflagellation occurring in functional male gametocytes and is measured by the formation of exflagellation centres or EC's. No ECs per  $10^4$  red blood cells were observed for the GNP lines in comparison to 34.85 ECs in the WT parental line 820 (gametocyte producer line) which served as controls (Figure 3-14).

### 3.5.4 Ookinete cultures

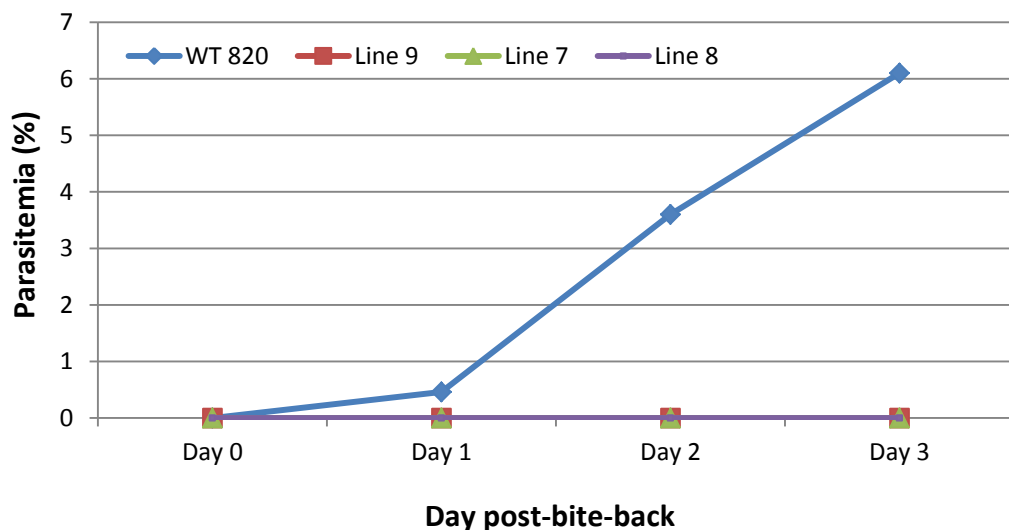
To further support the loss of gametocytogenesis in the three *de novo* GNP lines, overnight ookinete cultures (which assess the capacity of male and female gametocytes to fertilize and form zygotes which further develop to form ookinetes) were set up. The results show no observable ookinetes per 20 fields examined under Giemsa-stained smears from overnight ookinete cultures for each of the GNP lines studied in comparison to 19 ookinetes in the WT parental line 820 (gametocyte producer line) which served as positive control (Figure 3-14).



**Figure 3-14. Multiple bar chart demonstrating the absence of any detectable Exflagellation Centres & ookinetes.** ECs are expressed as number per  $10^4$  RBCs examined under a light microscope and ookinetes as number of ookinetes per 20 fields examined in Giemsa-stained smears in all the *de novo* GNP lines generated in the study ( $n=3$ ). This is in comparison to the WT line 820 for which an average 34.85 EC's were observed per  $10^4$  red blood cells examined and 19 ookinetes per 20 fields examined. The error bars denote  $\pm 2$  standard error around the mean.

### 3.5.5 Transmission through mosquitoes

As final conclusive evidence of the absence of any functional gametocytes in the GNP mutants, all the *de novo* GNP lines after negative selection and cloning (and the parental 820 line as a separate positive control) were passaged from the GNP-line infected mouse through mosquitoes and then back to a naïve mouse (mosquito transmission and bite-back experiments) to complete one parasite life-cycle. None of the mice from the bite-back experiment showed any parasites indicating that there were no gametocytes in the GNP lines when fed to the mosquitoes and hence no subsequent development of infection of the mosquitoes (Figure 3-15).



**Figure 3-15. Development of parasitemia from the mosquito bite-back (transmission) experiment in all the *de novo* GNP lines generated in the study in comparison to the WT line 820 as a positive control.** Clearly evident, the GNP parasites never grew in the mosquitoes as they totally lacked the sexual stages necessary to develop in the mosquito and maintain the life cycle of the parasite.

## 3.6 Characterization of the *de novo* GNP mutants

### 3.6.1 Asexual growth advantage

Asexual growth of the *de novo* generated GNP parasites appeared to be faster as compared to the parental gametocyte producer line and the

descendent “low” gametocyte producer lines. This was finally confirmed by mixing cloned and negatively selected GNP line 9 constitutively expressing CFP in a 1:1 ratio with wild type (PBANKA HP) producer line constitutively expressing RFP (Sinha et al., 2014; data not shown). Parasites were monitored daily by flow cytometry and after gating for infected cells the % of the population expressing either RFP (red), CFP (blue) or both (purple) reflecting mixed-multiply infected cells was calculated. The GNP parasites almost completely replaced the WT counterparts by day 11 post infection. These observations convincingly supported the hypothesis that the GNP phenotype offered a distinct growth advantage over their producer and “low” producer counterparts which helped the GNP population to eventually overgrow and replace the gametocyte producer parasites (see also the section 3.7 on mathematical modelling below).

### 3.6.2 Sexual growth

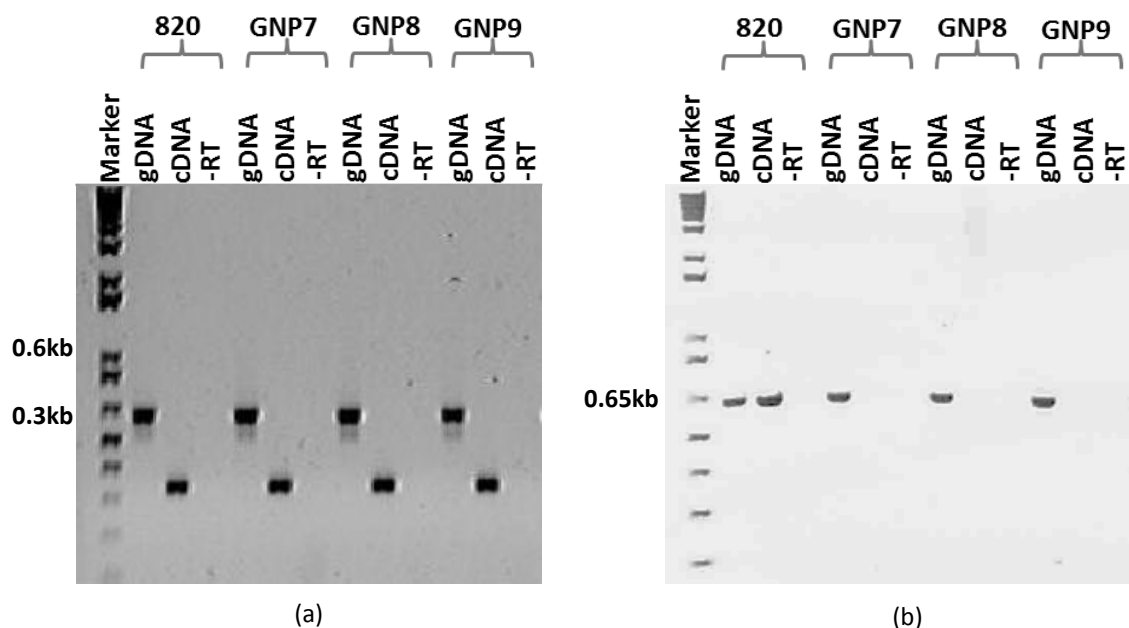
Further proof for loss of gametocytogenesis was established through examination of transcripts of the sexual-stage specific gene, *P28*, in the GNP lines using Reverse Transcriptase PCR (RT-PCR; Figure 3-16) and subsequently by northern analysis (Figure 3-17). The results showed no *P28* transcripts in all the generated GNP lines as compared to their isogenic wild type ancestor line 820. These findings in all the *de novo* GNP lines strongly support the absence of gametocytogenesis in malaria parasites.

#### 3.6.2.1 Absence of *P28* transcripts on RT-PCR

Presence of any sexual stage-specific transcripts for *P28* was ruled out using reverse-transcriptase PCR (RT-PCR). Briefly, ~4 µg of purified DNase treated total RNA from the WT parental line 820 and the three isogenic *de novo* GNP lines was subjected to RT-PCR (Section 2.2.6 of Materials & Methods).

The efficiency of the RT reaction (and hence contamination of the cDNA with the gDNA, if any) was tested by amplifying the constitutive *ALBA3* from all lines tested using primer sets GU515/GU516. Since *ALBA3* gDNA has an intron, its PCR amplification after RT reaction allowed the differentiation between amplicons from gDNA and cDNA on the basis of difference in expected sizes: 0.64

kb for gDNA and 0.32 kb for cDNA. The RT reaction yielded pure cDNA (uncontaminated with any gDNA) as evidenced by the presence of the smaller 0.32 kb band and absence of the gDNA band in ALBA3 RT-PCR amplicons (cDNA). Absence of the sexual stage specific *P28* cDNA band in the RT-PCR analyses from all three isogenic *de novo* GNP lines (in presence of the *P28* band in gDNA and cDNA from WT 820 line) further corroborated the findings of the northern analyses, strongly indicating an absence of sexual stages in the three *de novo* GNP lines. No bands in the -RT reactions ruled out the presence of any gDNA contamination in the samples (Figure 3-16).



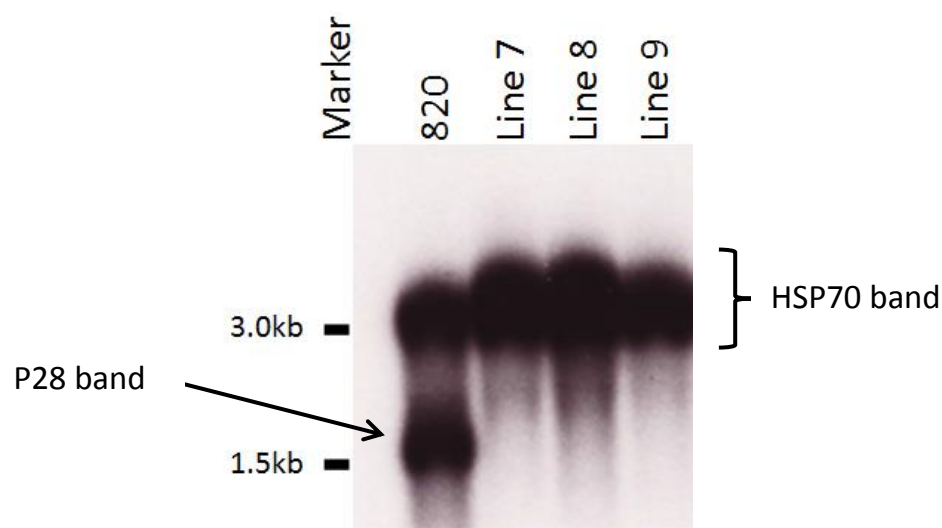
**Figure 3-16. Agarose gel electrophoresis pictures showing the amplified bands from RT-PCR reactions of *P. berghei* ALBA3 (constitutive) and *P. berghei* P28 (sexual stage-specific).** (a) shows the ALBA3 gDNA and ALBA3 cDNA (RT-PCR products) from the WT parental line 820 and the three isogenic *de novo* GNP lines to assess the efficiency of the reverse transcriptase reaction in terms of gDNA contamination of cDNA. Since the *alba3* gene (0.6 kb) contains an intron (0.3 kb) which reflects in the length of the *alba3* cDNA (0.3kb) thus providing a basis of detecting the gDNA (0.6 kb band) contamination of the cDNA. Absence of any 0.6 kb bands in all the samples rules out the possibility of the presence of any gDNA in the RT-PCR products. Sexual stage-specific *P28* was reverse-transcribed and PCR-amplified from the WT line 820 and all the three isogenic *de novo* GNP lines (b). Presence of *P28* gDNA bands in all the lines and *P28* cDNA only in the parental WT 820 line together with the similar findings from the northern analyses strongly establishes the absence of any expression of sexual stage-specific prototype *P28* from all the three isogenic *de novo* GNP lines thus documenting confirmed absence of sexual stages in the GNP lines – 7, 8 and 9.

### 3.6.2.2 Absence of *P28* transcripts on northern analysis

After ruling out the presence of any sexual stage forms of the parasite in all the *de novo* GNP mutants, the formation of sexual stage specific transcript



(*P28*) was also examined using the northern blot technique. Briefly, approximately 5 µg of RNA sample for each of the WT parental line 820 and the three GNP mutants was used for the analysis (for details see Section 2.2.7 of Materials & Methods). The absence of *P28* signals from all the *de novo* GNP mutants is consistent with the absence of mature sexual stages in these lines (Figure 3-17).



**Figure 3-17. Northern blot of total RNA from lines 820 (WT line) and the three naturally acquired *de novo* isogenic GNP lines – line 7, line 8 and line 9.** The blot was probed for sexual stage specific transcripts with  $^{32}\text{P}$  labelled *P28* dsDNA (~1.5 kb band in WT line 820 and no bands in the naturally acquired *de novo* GNP lines) and normalized with HSP70 probe (PBANKA\_071190; ~3 kb band in all lines). The absence of *P28* signal from the all the *de novo* generated lines is highly suggestive of absence of sexual stages in the GNP lines. (Note: the original blot had three extra lanes for showing the absence of *P28* signals in the three implicated gene knockout length variant lines after the lane for the GNP line 9; the same have been cropped out of the image (and shown elsewhere in Chapter 5, Figure 5-8) and the rest of the image merged to allow for the continuity of chapter and better understanding of the results discussed here).

### 3.6.3 Confirmation of the presence of the GFP/RFP expression cassette in the *de novo* GNP mutants

To rule out the possibility that the GFP/RFP expression cassette was neither deleted (in part or total) from the parasite genome, structurally rearranged in the genome or had become non-functional during the period of asexual multiplication thus giving rise to a false positive indication of loss of gametocytogenesis on FACS, the presence of full-length GFP and RFP expression

cassette in the original position (p230p locus of chromosome 3) was established by PCR in addition to the physical checks for the production of competent gametocytes that have already been detailed above. Evidence of the presence of GFP/RFP expression cassette was also sought from the whole genome re-sequencing data which also suggested the retention of the expression cassette in the GNP mutants, thus confirming that the FACS findings and biological observations were, in fact, correct.

### 3.6.3.1 PCR confirmation of the expression cassette

PCRs were carried out to amplify and thus confirm the presence of full length GFP and RFP open reading frames along with amplification of the 3' integration PCR product using appropriate oligonucleotide pairs (Table 3-1 and Figure 3-18a). The presence of appropriately sized bands for GFP and RFP (Figure 3-18b) indicates the presence of unaltered GFP/RFP expression cassette in the original genomic locus (p230p locus of chromosome 3).

Table 3-1

Oligo set	Details	Expected band size with DNA (kb)				Ta (°C)	Ext. Time (min)
		820	GNP7-9	Water	HP		
1034/1032	GFP	0.66	0.66	--	NA	53	5
1030/1031	RFP	0.68	0.68	--	NA	53	5
1030/1029	3' Integration Check	2.61	2.61	--	NA	53	5
13/14	Control (HP)	NA	NA	NA	0.6	53	5

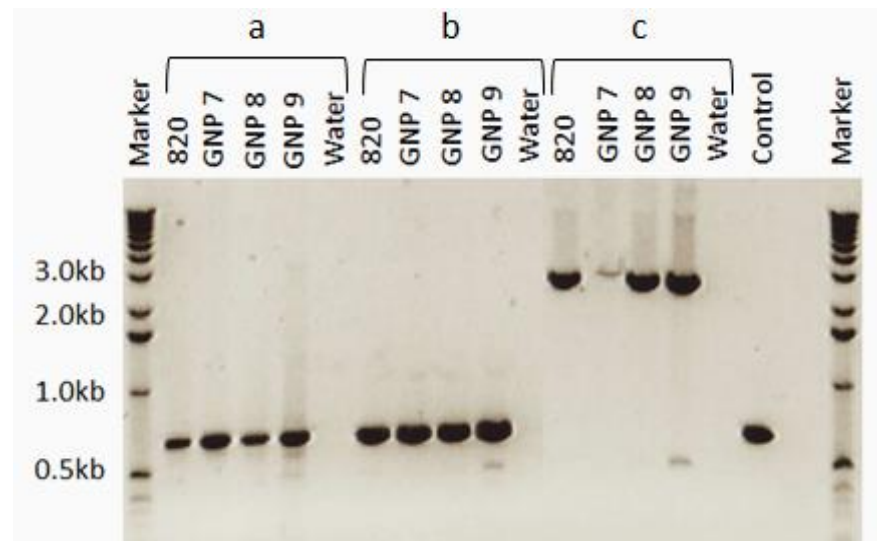
**Table 3-1. Basic PCR reaction settings** for amplifying various regions of and around the GFP/RFP expression cassette at the 230p genomic locus to identify the presence of the expression cassette in the genome of the GNP lines.



(a)

**Figure 3-18 (a). Schematic (not to scale) showing the genomic 230p locus (blue thin bars) on *P. berghei* chromosome 3.** Also shown is the presence of a positive-negative drug selection marker cassette (aqua coloured thick bar) and GFP/RFP expression cassette containing *e-GFP* gene (green thick bar) driven by a 5' male-specific promoter of *dynein heavy chain* (olive green thick bar) and 3' UTR of *Pb calmodulin* gene (purple thick bar). The RFP (red thick bar) is under a 5' female-specific promoter of *ccp2* (light red thick bar) and *p48/45* 3'UTR (light orange thick bar). The primer sets used for amplifying and the expected sizes of the amplicons for the regions "a" (forward

and reverse from eGFP), “b” (forward and reverse from RFP) and “c” (forward from RFP and reverse from outside the 230p locus) include GU1034/GU1032 (0.66 kb), GU1030/GU1031 (0.68 kb) and GU1030/GU1029 (2.61 kb), respectively.



(b)

**Figure 3-18 (b).** PCR showing the amplification of appropriate sized bands for the WT line 820 and the GNP lines 7, 8 and 9 denoting that the GFP/RFP expression cassette is retained in place by all the natural GNP mutants generated *de novo* in this study.

This further implies that the absence of GFP/RFP signals for the GNP lines 7, 8 and 9 as reported by FACS was real and not simply due to alteration and/or absence of the expression cassette itself and thus indicates non-production of gametocytes.

### 3.6.3.2 Evidence from whole genome re-sequencing data

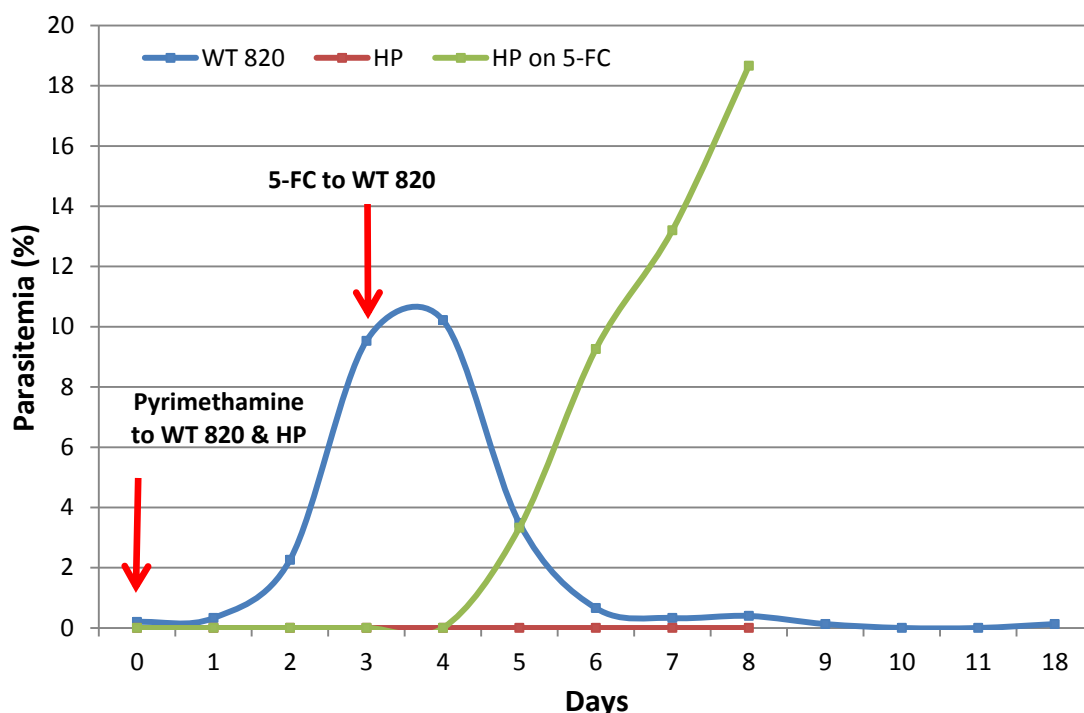
Re-sequencing the GNP lines and mapping on the gametocyte producer ancestor lines presented a specific problem. Line 820cl1m1cl1cl2 (cloned, parent line of the GNP mutants) was initially mapped onto the reference line ANKA which does not possess the GFP/RFP expression cassette and thus the cassette was assembled in fragments on chromosome 3 and places elsewhere (chromosomes 4, 7, 10, 11, 13 and bin) as expected due to mapping on to the promoters used to drive the GFP/RFP expression. So when the GNP lines were sequenced and mapped on to 820, the same observations regarding the presence of the cassette were made.

### 3.6.4 Selectable Marker Cassette and *de novo* GNP mutants

After the experiment to *de novo* generate the isogenic gametocyte non-producer mutants from the parental 820 line was mid-way through, in an independent experiment, the growth kinetics of the parental 820 line (820cl1m1cl1cl2) under pyrimethamine selection (see Section 2.1.5 of Materials and Methods) suggested that the parasites were in fact resistant to the drug (Figure 3-19). The WT parental line 820 and a non-fluorescent labelled and non-drug selectable marker bearing line HP (as negative control) were grown under pyrimethamine drug pressure from day 0 of the infections. As evident from the figure, the parasites grew normally in the WT line 820 as these were supposed to be resistant to the drug in contrast to the HP infection which were controlled by the drug. This strongly suggested that the WT line 820 still possessed the dhfr gene in the drug selectable marker cassette (SMC; despite indication that a round of negative selection was previously applied to recycle the SMC).

This was further confirmed when the drug 5-FC (1 mg per mouse per day; intra-peritoneally) was added to the WT line 820 regime from day 3 post-infection. The retention of the drug SMC (a positive-negative SMC) would confer sensitivity to the lethal drug 5-FC through *yfcu* - the rapid decline in parasitemia following 5-FC administration supported the presence of the drug SMC in the WT parental line 820. The absence of sensitivity to 5-FC in the HP line (which does not possess the SMC) was also demonstrated by growing the new line HP in presence of the drug 5-FC - the parasites grew normally signifying the absence of the SMC.

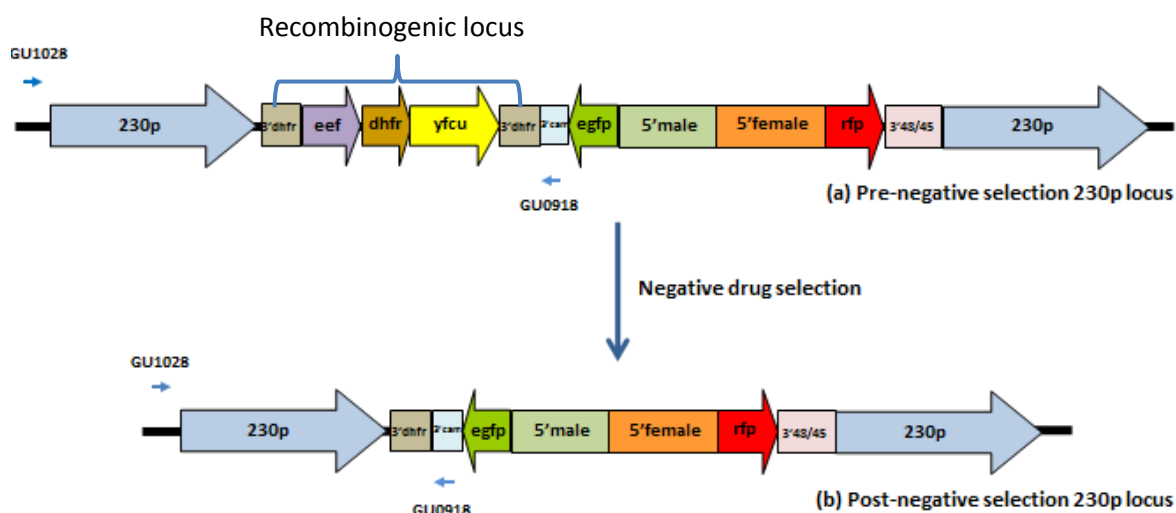
Thus the experiment implied that the WT parental line 820cl1m1cl1cl2 still possessed the positive-negative SM cassette integrated into the genome suggesting that the SMC was not recycled after a round of negative selection. This was later confirmed by PCR analysis. Since the *de novo* GNP lines 9, 7, and 8 were isogenically derived from the WT parental line 820, these GNP lines were also suspected to still possess the drug SMC. This implied that any further genetic manipulation of these GNP lines would not be effective in the future as the SMC was not recycled, thus demanding a negative selection on these lines followed by cloning.



**Figure 3-19. Drug assay showing the retention of the SMC.** The assay involved growing the WT parental line 820 and non-fluorescent high gametocyte producer line HP under pyrimethamine selection in drinking water and 5-FC selection intraperitoneally to show the retention of the positive-negative drug SMC in the 820 genome. Pyrimethamine (oral) was started from day0 to both WT parental line 820 and HP. 5-FC (i.p.) was added to the WT parental line 820 on day3. Independently, 5-FC was also administered to the HP line alone. The kinetics of all the lines strongly suggested the retention of the drug SMC by the WT parental line 820 and thus also by all three *de novo* GNP lines derived from 820.

#### 3.6.4.1 Removal of the drug selectable marker cassette (SMC) from the GNP lines

In order to be able to perform any subsequent complementation experiments on the GNP lines, the SM cassette must be eliminated from the genome of the GNPs. Because of the relative paucity of sufficient number of drug selectable markers for *Plasmodium* genetic manipulations (Ganesan et al., 2011; Lin et al., 2011), SM recycling provides an alternative way to reuse the same/limited drug selective markers for any number of successive experiments requiring these drugs to select the mutants (Braks et al., 2006; Maier et al., 2006; Orr et al., 2012). This was achieved by putting these lines on the drug 5-FC (Negative selection; 1.5 mg/ml) added to drinking water and administered for a period of approximately 8-10 days, depending upon the experiments. A schematic showing the negative selection based removal of the drug SMC from the GNP mutants is shown in Figure 3-20.



**Figure 3-20. A schematic (not-to-scale) of the *P. berghei* 230p locus before (a) and after (b) negative selection process using 1.5 mg/ml of the drug 5-FC in drinking water. Also shown is the position of the oligonucleotide pair (GU1028; GU0918) used for amplifying the full length of the SMC in the GNP mutants alongside the WT line 820.**

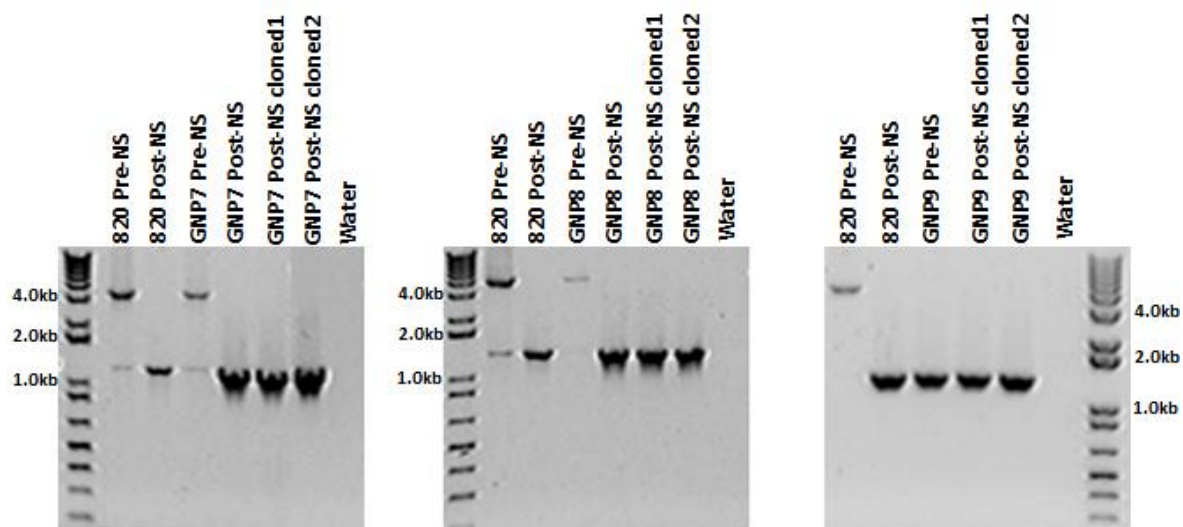
Negative selection of the GNP line 7 produced 4 negatively selected lines from GNP m7 and hence they are called m7m1 to m7m4. Similarly, GNP line 8 & GNP line 9 produced 2 negatively selected lines, namely m8m1-2 and m9m1-2.

### 3.6.4.2 Cloning of the negatively selected GNP lines

Having removed the SM cassette (SMC) from the GNP genomes through marker recycling under negative selection, these SM-minus GNP lines (particularly m7m4, m8m1 and m9m1 were cloned by the limiting dilution method. It was necessary to clone the negatively selected line before attempting any further genetic manipulations on the line. This is of paramount importance when the intended planned experiment is complementation to restore the gametocyte producer WT phenotype.

The three independent cloning experiments resulted in two PCR-confirmed clones each from GNP7 (m7m4cl1-2), GNP8 (m8m1cl1-2) and GNP9 (m9m1cl1-2) as shown in Figure 3-21. Equal amount of purified DNA (~100 ng) from each line was used for the PCR conformation using the oligonucleotide pair (GU1028/GU0918; Figure 3-20) for amplifying through the full length of the SMC in the GNP mutants before and after negative selections (pre-NS and post-NS,

respectively) alongside the WT line 820. The expected band-size for pre-NS lines was 4.1 kb and for the post-NS cloned lines was 1.3 kb.



**Figure 3-21. Agarose gel electrophoresis of the PCR-amplified DNA before and after negative selection.** Pre-negative selection (pre-NS) and post-negative selection (post-NS; cloned and uncloned) for SMC recycling was applied on the WT line 820 and GNP lines 7, 8 and 9 using the oligos GU1028/GU0918. Expected band size of 4.1 kb was observed for all the pre-NS samples signifying the presence of the SM cassette. A desired band of 1.3 kb on all the post-NS samples suggests the elimination of the SM cassette from the GNP lines.

### 3.7 *Evidence from mathematical modelling of the kinetics*

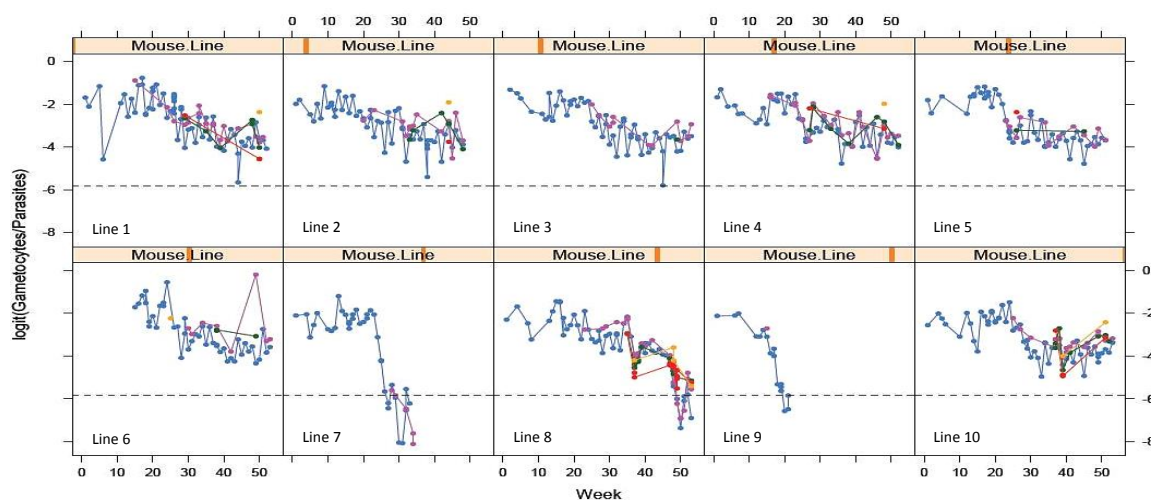
Mathematical modelling (interim analyses) of the kinetics of parasitemia and gametocytemia in all the ten isogenic lines was performed by Dr Mathew Denwood and Prof Dan Haydon, University of Glasgow. Although the dataset given was complex, the main parameter that was included was the change in gametocytemia (percent) over time as a function of repeated mechanical passages through mice.

#### 3.7.1 *The distribution of the data for the controls*

The available data for the controls (representing the background noise) closely followed a normal distribution after logistic transformation of the data which was done in accordance with the exponential pattern of parasite growth. The mean background noise was estimated to be approximately 3 false positive



cells per 1000 cells examined with a standard deviation ranging between 0.8 and 1.7 (Figure 3-22).

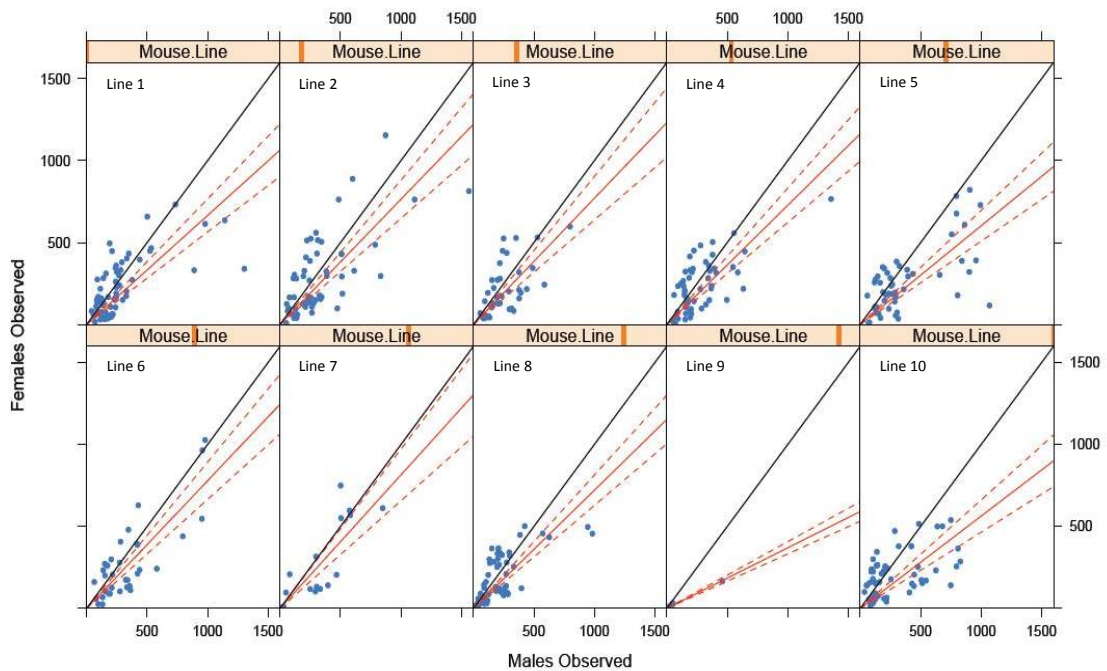


**Figure 3-22: Plotting of the raw data after a logistic transformation in the ten isogenic lines.** The data also revealed a linear trend for most mouse lines, which is reassuring (see graph above – days are shown in different colours & the 'random noise' baseline is shown dashed). The overall picture strongly suggested overall decreasing gametocytemia kinetics in all the lines. In the eventual GNP lines, GNP lines 7, 8 & 9, the lines representing the gametocyte kinetics crosses and dips below the 'random noise' baseline.

### 3.7.2 Sex Ratio

It was found from the data that the proportion of female gametocytes was consistently lower than the male counterparts for all ten lines (Figure 3-23). In general, female biased sex ratios are seen in *Plasmodium* in nature (Read et al., 1992, 2002; Robert et al., 1996; Sowumni et al; 2008; West et al, 2001), but nevertheless, neutral and male biased sex ratios have also been observed (Read et al., 1992 and 2002; Robert et al., 1996). Briefly, sex ratios in *Plasmodium* are very dynamic and may involve a range including female- and male-biased ratios (Teboh-Ewungkem and Wang, 2013). Although *P. berghei* follow the same principles of sex ratio as any other *Plasmodium*, a male biased sex ratio has been observed after phenylhydrazine treatment of the mouse (Chris Janse, Unpublished observations, Leiden University Medical Centre). Relatively lower females may also be due to weaker expression/detection of RFP in comparison to GFP.



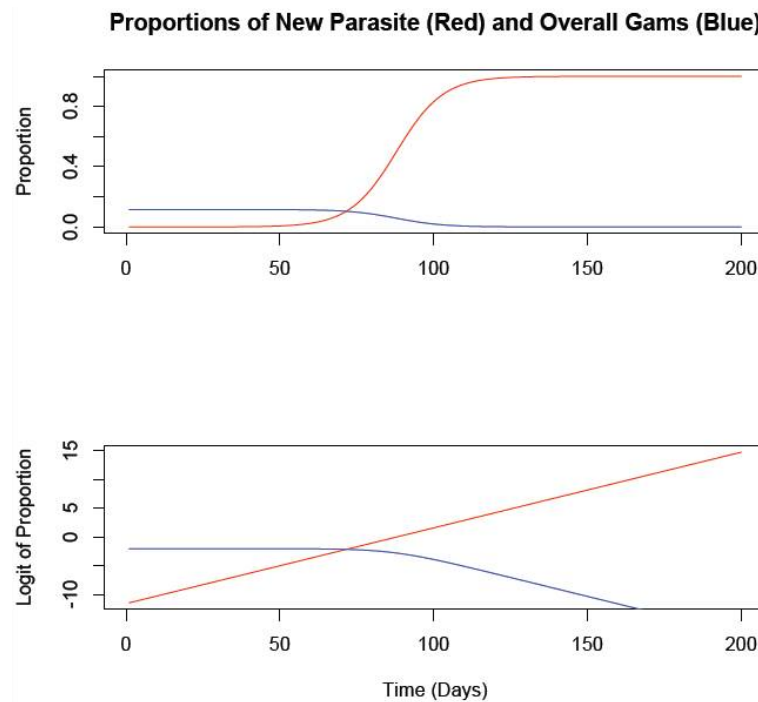


**Figure 3-23: Statistically modelled female (vertical axis) and male (horizontal axis) gametocyte kinetics of the ten isogenic lines.** The proportion of female gametocytes (red straight line) almost always appeared to be lower than their male counterparts (black straight line). The variance around the proportion of female gametocyte (red dashed lines) was also always below the black line for males.

### 3.7.3 *Predicting post-mutational timescale changes in parasite population within a mouse*

The available data was also be used to predict the time needed for the gametocytemia to drop to zero following the generation of a GNP mutation in a single parasite. For example, if such a GNP conferring mutation arose in an individual parasite reducing gametocyte (%) production from 11% (the average at week 1 of the study) to 0%, the time the mutated parasite would take to increase from one to 7000 (this number is based on the average parasitemia at week 1 corrected for the circulating blood volume of the mouse) would be approximately 150 days (Figure 3-24), assuming that:

- (1) An asexual parasite takes 24 hours to produce daughter cells
- (2) An average of 16 daughter cells per parent is produced
- (3) There is no competitive advantage incurred from the genetic mutation except that which is directly attributable to the reduction in gametocyte production (i.e. death rates and division time remain the same)



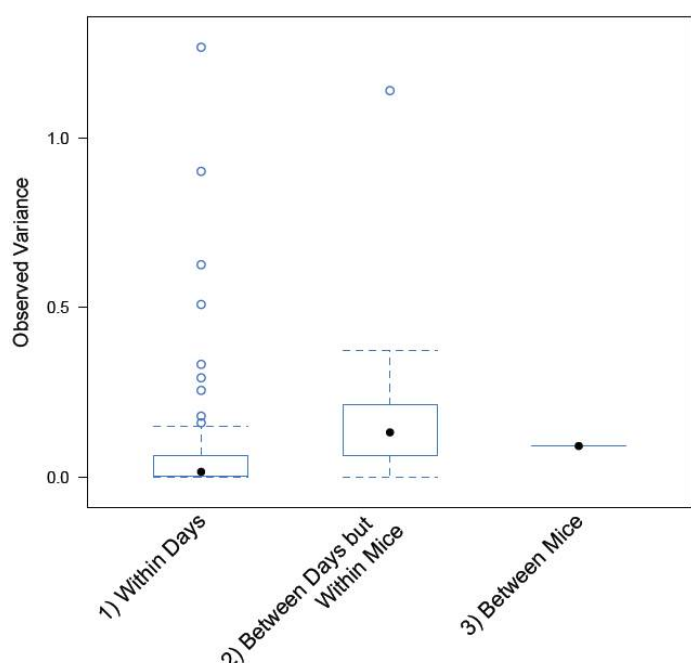
**Figure 3-24. Change in gametocytemia kinetics due to the presence of even one parasite harbouring the GNP mutation.** The upper graph displays the raw (untransformed) data whereas the bottom graph shows that the proportion of parasites with the mutation (red) increases linearly on the logistic scale, but this is manifest as a long lag period before an appreciable change (decrease) in the average gametocyte production of the overall parasite population (blue). Also shown are the approximate number of days for the total gametocytemia to reach the 0% level as the GNP-mutation bearing parasite population grows from 0% to 100%.

In addition to the estimate for the time needed for a single GNP genetically coded parasite to overgrow and fully replace the gametocyte producing WT parasites, independent growth competition assays involving a mix of GNP and WT population demonstrated the growth advantage offered by the GNP parasites to overgrow and eventually replace their WT counterparts (Sinha et al., 2014; data not shown).

### 3.7.4 Variance Partitioning

Analysis of between sample variability was highly convincing not only in terms of the level of confidence that could be placed in a single observed gametocyte (%) observation but also for planning future similar studies with optimum efficiency. Conceptually, the variation in the study can be partitioned into the following (Figure 3-25):

- Between mice variability: Variation between samples taken from different mice infected with the same parasite
- Between days variability: Variation between blood samples taken (over a short time period) from the same mouse [There was no evidence that these are anything other than random over a period of 5 days]
- Within days' variability: Variation between 100 $\mu$ l sub-samples taken from the same blood sample
- Variation due to the Binomial distribution (counting error) and random 'noise' of the lab process



**Figure 3-25. Boxplots showing the different possible sources of variability within the experimental setting.** As seen, the variability between mice, although based on a single observation, is very low as compared to the variability within days of experiment (variation between sub-samples from the same blood) and variability between days but within mice (blood samples taken over few days from the same mouse) which was the greatest.

The above analyses on variation suggested that the variation between days appeared to be greater than that within days but the variation between mice may not be more than that between samples from the same mouse (Denwood et al., 2012). Although it would not be statistically appropriate to place too much confidence in the single observation, it could be inferred that variation between mice did not contribute a substantial amount of variation.

### 3.8 Summary and discussion

Generation of at least one *de novo* gametocyte non-producer (GNP) line (out of the potential ten passaged lines) from the cloned high gametocyte producer line 820 was the first major objective of the study. Within the period of approximately one year, ten parallel but independent lines (line 1 to 10) derived from a common ancestral line (PBANKA 820cl1m1cl1cl2) were serially passaged from one mouse to another until the end of the study period or development of a GNP phenotype, whichever was earlier. At the end of the study period, 3 out of 10 potential lines turned out to be GNP lines on flow cytometry and Giemsa-stained smear examination.

These *de novo* GNP lines were later confirmed for absence of viable gametocytes directly by *in vitro* (exflagellation assay and ookinete counts) and *in vivo* (passaging through mosquitoes) experiments and indirectly by showing absence of sexual stage specific transcript of *P28* (Northern and RT-PCR analyses). Upon confirmation, these *de novo* GNP lines were subsequently negatively selected, cloned and cryopreserved for further characterizations. A brief summary of details of passages involving each of the 10 lines is shown in the following table (Table 3-2). The rest of the passaged lines were characterised as low gametocyte producers at the end of the study period. These lines remain intriguing in terms of the revelations of genotypic changes associated with the phenotypic transition from high gametocyte production to no gametocyte production.

In general and simply by visual inspection, the gametocytemia trend could be broadly categorised following a three phase pattern over the study period of ~52 weeks in all the isogenic serially passaged lines, barring the GNP line 8 which showed somewhat different kinetics (exemplified by the kinetics in the GNP line 3; Figure 3-26a). For all the other lines, the gametocytemia fluctuated between 7% and 17% initially during the first half of the total weekly passage period (herein termed as the normal fluctuating phase). The second phase of gametocytemia trend could be inferred as the gradual declining phase where the percentage of gametocytes gradually declined over the next 10 week period to ~3%. The first two phases were almost similar in all the lines (except line 8) but

the third phase differed for most lines between those that ultimately became GNP and those that remained (low) gametocyte producers after ~52 weeks of mechanical passage.

Table 3-2

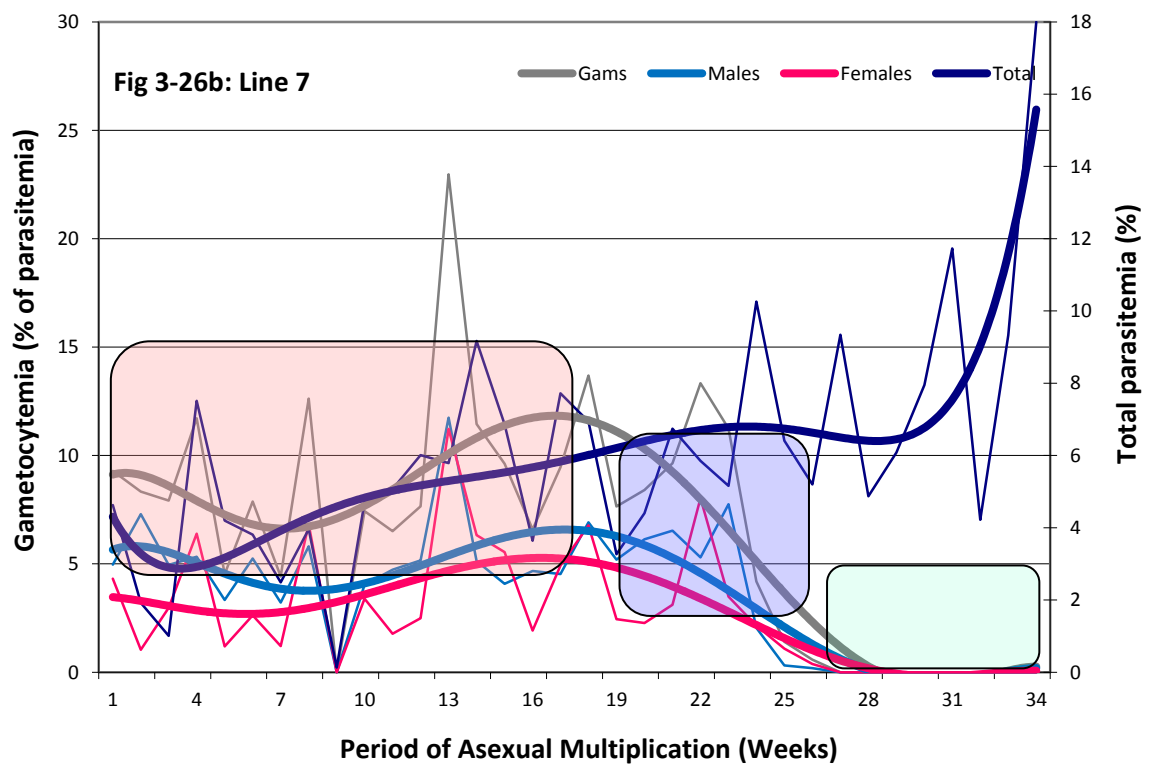
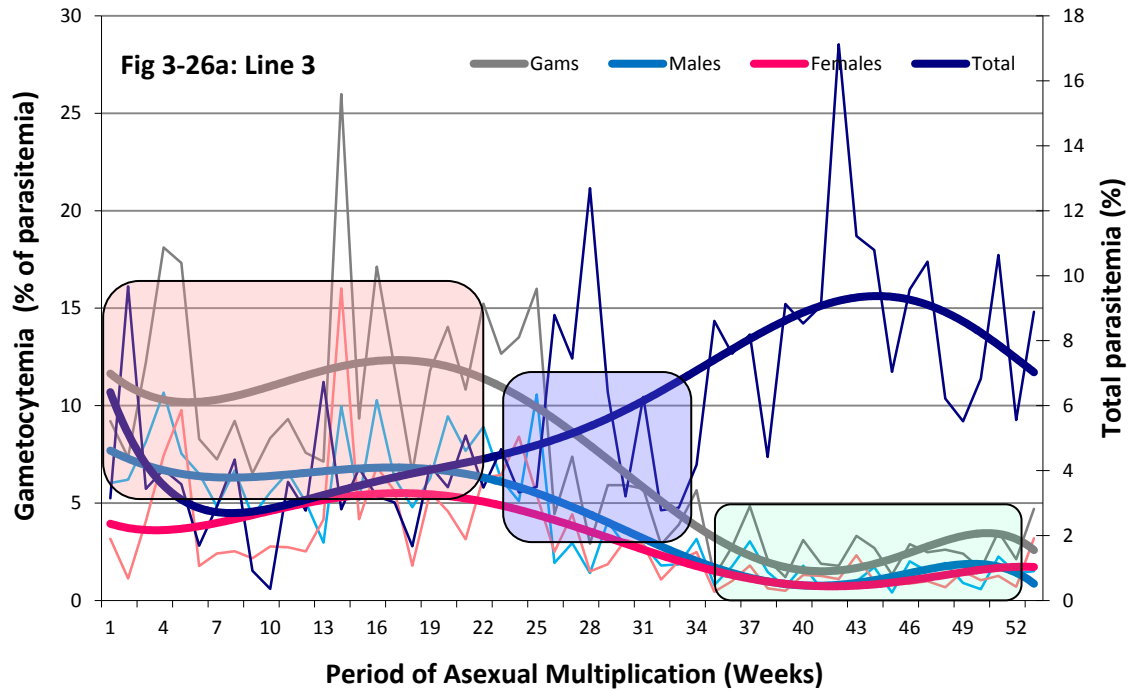
Line	Total number of weekly passages	Number of times reactivated from stabilates	End point	End-point phenotype
1	52	4	End of study	Low producer
2	47	5	End of study	Low producer
3	53	1	End of study	Low producer
4	51	6	End of study	Low producer
5	50	7	End of study	Low producer
6	53	3	End of study	Low producer
7	32	5	GNP	GNP
8	53	5	GNP	GNP
9	21	5	GNP	GNP
10	53	3	End of study	Low producer

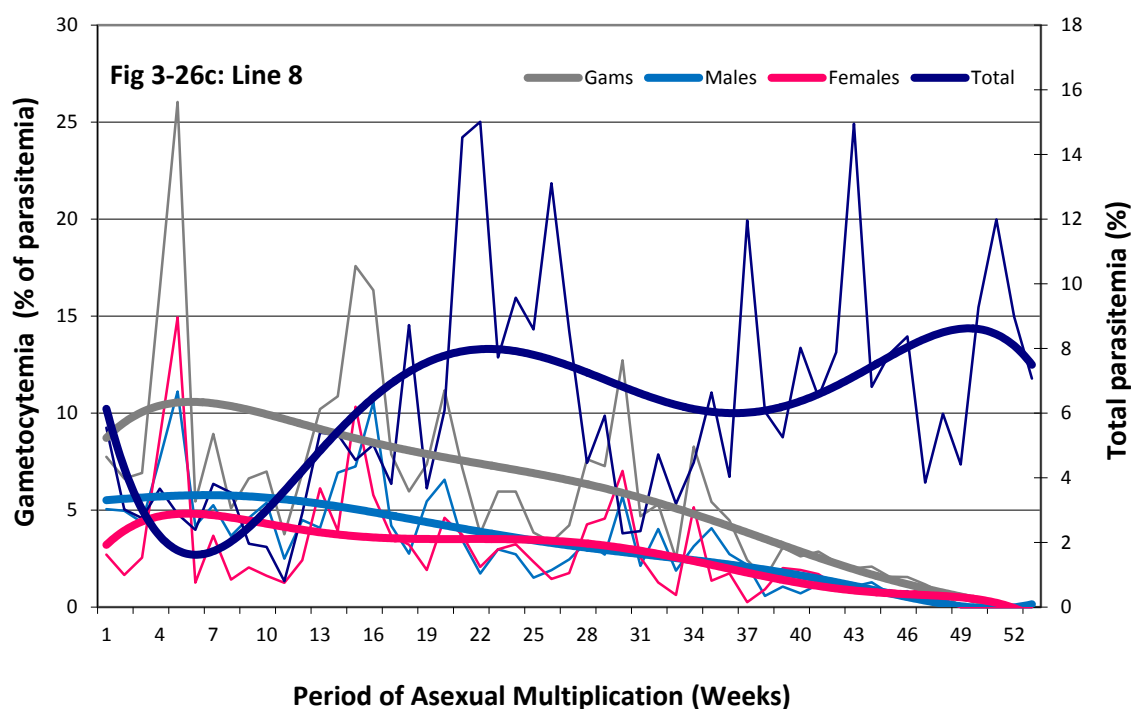
**Table 3-2. Summary of weekly passages of the ten isogenic lineages derived from the common WT ancestor line 820.** Clearly evident is the generation of three GNP lines at the end of the study period (lines 7-9). These lines had phenotypically confirmed GNP status each at weeks 32, 21 and 53 of the serial passages. Lines 1, 3, 6 and 10 were passaged for at least 52 weeks whereas lines 2, 4 and 5 could be passaged only for 47, 51 and 50 weeks due to the frequent failure of weekly infections (indicated by the number of times the lines had to be reactivated from the cryo-preserved corresponding stabilates. The lines which did not end up being GNP had a lower (than starting) parasitemia which could be intriguing to explore.

In the lines which did eventually develop as GNP lines barring line 8, this third phase showed a continued decline (as an extension to the declining phase) in gametocytemia until eventually no gametocytes could be identified (exemplified by kinetics in the GNP line 7; Figure 3-26b). In the rest of the lines, the third phase again showed a fluctuation in low level gametocytemia (herein termed as the low fluctuating phase), but generally at levels lower than the initial gametocytemia being, between 1% and 3%.

The kinetics of GNP line 8 exhibited a unique pattern (Figure 3-26c) wherein the gametocytemia fell gradually, at almost a steady rate, since the start of period of absolute asexual multiplication, maintains the same trend throughout till the line no longer produced gametocytes at the end of 52 weeks of asexual multiplication. The difference in the kinetics between line 8 and the rest of the lines opened an intriguing debate which might underpin a molecular

mechanism mediated by the necessity for a cascade of mutants before a line becomes a GNP line.





**Figure 3-26: Representative growth kinetics from a gametocyte producer line (line 3, fig 3-27b); GNP line (line 7, fig 3-27b), and another GNP line (line 8, fig 3-27c).** X-axis shows the period of asexual multiplication (weekly passage number) in weeks. Total parasitemia (shown as navy blue trend line, as %age of infected red blood cells out of 1500 red blood cells examined) is represented along the secondary Y-axis whereas gametocytemia (shown as grey trend line, as %age of total gametocytes, males and females, out of the infected red blood cells) is represented on the primary Y-axis. Total percent of gametocytes were further sub-categorised into percent male (blue line) and percent female (pink line) gametocytes out of the total infected red blood cells. Each individual weekly observation is joined by a spiky line of representative colour whereas the smoothed line is the best-fit polynomial trend line that best explains the trend of fluctuations of individual weekly observations. The pink, blue and green semi-transparent rectangles represent the hypothesised phases 1 to 3, respectively of the observed gametocytemia kinetics. The tails (pink and blue) observed for gametocytemia in line 7 (Figure 3-26b) after passage week 31 is due to the smoothed polynomial being fitted for the kinetics and do not represent actual gametocytemia.

The recent identification of many genes essential for gametocytogenesis in *P. falciparum* through transposon mediated mutagenesis also suggests that these are essential for gametocyte maturation (Ikadai et al., 2013). Although the systematic gene deletion (*P. berghei*) and transposon mutagenesis (*P. falciparum*) have identified genes essential for gametocytogenesis, reverse genetics screens have not indicated an absolute GNP phenotype in any of these mutants to date.

### 3.9 Future directions

Is there a possibility of a reversal of GNP mutant phenotype to WT in nature or *in vitro* is an interesting question to be addressed in future. There

exists limited evidence of gametocytogenesis being induced in lines which naturally evolved to the GNP phenotype as a result of continuous maintenance in the absence of mosquito bites by adding chemicals such as ammonium compounds and Berenil (Ono and Nakabayashi, 1990; Ono et al., 1993) or by adding RBC lysate (Carter and Miller, 1979) or mammalian hormones (Maswoswe et al., 1985; Lingnau et al., 1993) to *P. falciparum* culture. In terms of evolutionary advantage, investment in sexual stages would be inefficient in the absence of uptake by the vector but some natural cues (like onset of transmission season and bites from uninfected mosquitoes) associated with increased gametocytogenesis have been mentioned (Ouedraogo et al., 2008; Paul et al., 2004; van der Kolk et al., 2003).

It would also be worthwhile to statistically predict from a mathematical model, the approximate time (in terms of weekly passage number) when the currently cryopreserved low gametocyte producer lines might develop the GNP phenotype and which lines should be preferentially chosen on the basis of the least distance/time to become a GNP in terms of the number of passages needed. Such a model could also be fine-tuned to predict the period of latency from the time of precise occurrence of a causal mutation and the generation of the first detectable GNP phenotype in such lines.

Further, accumulating recent evidence in support of the involvement of cell-to-cell signalling in malaria parasites might be one of the possible triggers to commitment of the parasite to sexual development mediated by microvesicles and exosome-like vesicles (Mantel et al., 2013; Regev-Rudzki et al., 2013). In this regard, the comparative analyses of the metabolomes from the cryopreserved plasma of the *de novo* generated GNP line/s and its immediate gametocyte producing ancestor/s could help us identify if there are any molecular signatures of possible communication/s leading to the generation of the GNP phenotype and possibly decode the language used by parasites for such communication (del Portillo and Chitnis, 2013).



**4 Whole genome re-sequencing, assembly and genome-wide comparison of *de novo* GNP mutants and the isogenic parental line**

## 4.1 Introduction

One of the most significant achievements of next generation sequencing technologies is the ability to generate a tremendous amount of reliable data cheaply and rapidly (Metzker 2010). These latest and improved technologies (including Roche/454, Illumina/Solexa, Life/APG and Helicos BioSciences) dramatically increase the sequencing throughput by laying millions of fragmented/sheared DNA molecule on a single solid support and sequencing all these fragments in parallel, the so called Massively Parallel Sequencing (Pop and Salzberg, 2008; Li et al., 2009; Tucker et al., 2009;).

Whole genome re-sequencing or sequencing the genomes for which there exists a reference genome has opened the ways to strongly establish associations between the documented phenotype and the genotype (Stratton, 2008). Application of whole genome re-sequencing approaches in *Plasmodium* has helped narrow the number of candidate mutations implicated for a particular phenotype. It has now been possible to precisely identify, locate and study the exact nature of the genetic mutation/s behind various important phenotypes generated by applying forward genetic screens in the isogenic lines of different *Plasmodium* species (Modrzynska et al., 2012). These include identification of specific genetic mutations responsible for generating resistance to common anti-malarial drugs like Artemisinin (Hunt et al., 2007; Hunt et al., 2010), Sulphadoxine (Martinelli et al., 2011), or Mefloquine and Lumifantrine (Borges et al., 2011) and Chloroquine (Modrzynska et al., 2012) in the model (rodent) malaria parasite, *P. chabaudi*.

Following the same principles, a whole genome re-sequencing approach was applied here by mapping the whole genome sequencing reads of the three independent naturally selected isogenic gametocyte non-producer (GNP) lines (generated by repeated in vivo passage; Chapter 3) on the *de novo* assembled reference genomic sequence of their parental wild type line (WT 820), and also on the already assembled reference genome of *P. berghei* ANKA, to understand the precise molecular basis of gametocytogenesis in malaria parasites. To obtain supporting evidence, the mapping was also performed from the reads of two independent pre-existing gametocyte non-producer lines: clone 1 of the K173

lineage (Vincke and Lips, 1948; Janse et al., 1989; Janse et al., 1992) and clone 233L derived from the parental clone 234L of the ANKA lineage (Killick-Kendrick, 1974; Dearsly et al., 1990).

A flow diagram highlighting the major procedures performed from preparation of the DNA for sequencing up to the analysis of identified mutations is shown in Materials & Methods (Figure 2-2). A summary of the basic algorithms used in generating a reference assembly, mapping and variant calling is presented in Table 2-1 of Materials & Methods.

## 4.2 Illumina/Solexa whole genome sequencing and assembly of the GNP and WT isogenic parental lines

This section was entirely performed by Dr Thomas Dan Otto and Mandy Sanders and colleagues at the WTSI. All the *de novo* GNP isogenic lines (lines m7, m8 and m9) along with their common parental line (820cl1m1cl1cl2) were sequenced by whole genome sequencing on the Illumina/Solexa sequencing platform using paired-end reads. In addition to this, a pre-existing GNP line K173cl1 and line 233 along with its gametocyte producer parent, line 234, were also sequenced. A summary of reads for each project including accession codes are given in Table 4-1. Data are available at <http://www.ebi.ac.uk/ena/data/view/ERP000253>.

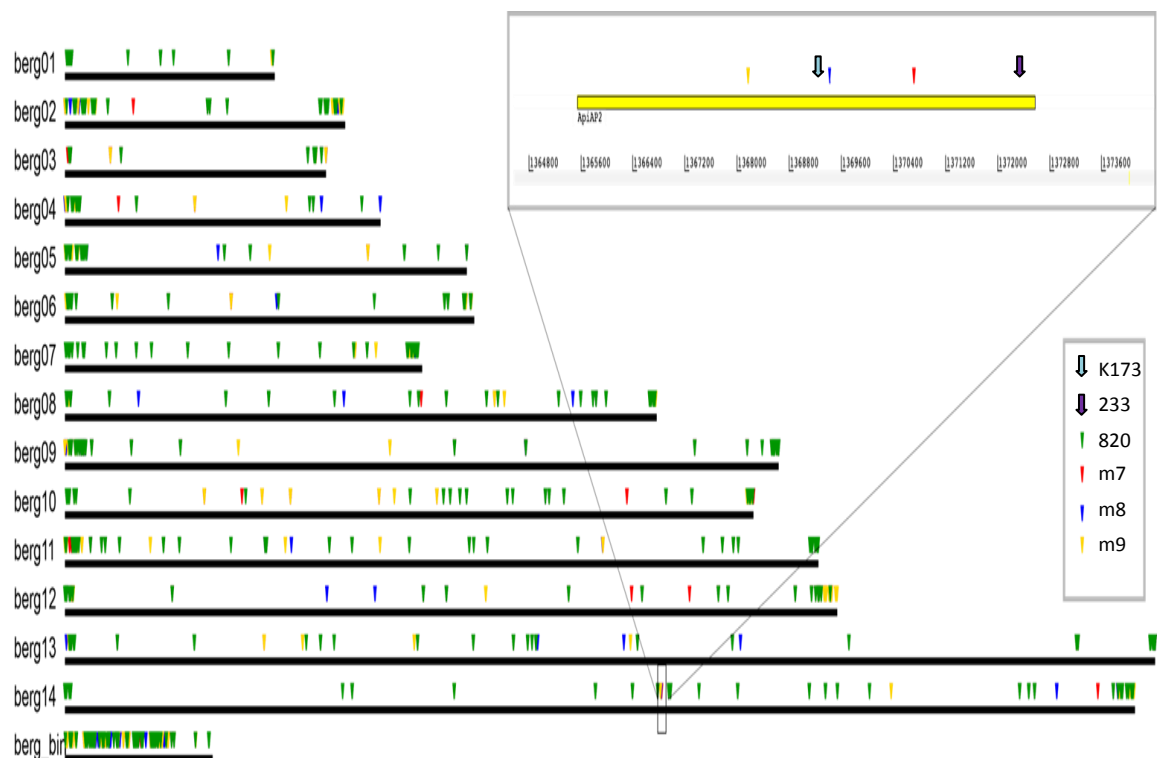
Table 4-1

Isolate / clone	Number of reads	Accession number	% mapped reads on <i>P. berghei</i> ANKA 820	approximate coverage x
820	52,888,156	ERS012301	90.8	194
m7	66,824,730	ERS016741	91.77	248
m8	123,118,738	ERP000253	90.43	450
m9	63,297,465	ERS012302	87.5	224
233	29,330,944	ERS004442	89.23	106

Table 4-1. Parameters of WGS of GNP *P. berghei* lines used in this study

### 4.3 Mutations identified in the study lines after mapping them on the reference assembly lines

After mapping onto their corresponding reference genomes, the mutations (SNPs and indels) were called using SSAHA2 and SAMtools algorithms (see Materials & Methods section 2.6.3). Two separate lists of SNPs and indels were created (Tables 4-3 and 4-4) based upon (a) mapping of all lines onto 820cl1m1cl1cl2 reference assembly; and (b) mapping of all lines onto ANKA reference assembly. A list was also generated for mutations in the un-translated regions (UTRs) of the genes - within 1 kb upstream (5'UTR) and 1 kb downstream (3'UTR) of the coding sequences, including those in the coding sequences (Table 4-5). The WGS and SNP analytical pipeline is summarized in Figure 4-1.



**Figure 4-1. Whole genome sequencing and SNP calling analyses.** A *de novo* assembly of the parental WT line 820 was created. All the three natural mutant lines were also sequenced and assembled using the 820 assembly as the reference. SNPs were called by mapping the GNP lines on to the 820 parental line for all the 13 chromosomes. Also, the WT parental line 820 was mapped on itself to identify any sequencing and/or base calling errors. The mutations for each line (inverted triangles) are denoted by green (for 820 line; possible errors), red (for GNP line 7; probably real), blue (for GNP line 8; probably real) and yellow (for GNP line 9; probably real). The inset shows the gene PBANKA\_143750 on chromosome 14 with unique mutations in all the three natural GNP mutants.

### 4.3.1 ***Mutations identified within the open reading frames in the GNP study lines after mapping onto the 820 reference assembly***

A total of 4352 SNPs and 184 indels were detected when all the lines (820cl1m1cl1cl2, ANKA, 233, 234, line 7, 8 and 9) were mapped onto the 820 reference genome. All mutations were first filtered using two or three subsequent filtering steps to segregate the suspected erroneous and/or lineage specific mutations from the possibly real causative and thus analysable mutations. The general schema for mapping and filtering mutations is shown in Table 4-2.

Candidates were first filtered for mutations arising out of obvious artefacts and possible errors in mutation calling programs. Those mutations were considered erroneous which were called by mapping the WT line 820 sequencing reads on its own *de novo* assembly and were identified in all the other lines. Such mutations were eliminated (as errors in mutation calling) not only from those detected in 820 but from those identified in other lines as well (FILTER 1).

Table 4-2

Reference	Mapped	Interpretation of mutations	Action
820	820	False positive (errors)	Filtered
	ANKA	Strain difference	Filtered
	234	Strain difference	Filtered
	GNP (233, m7, m8 & m9)	<b>Probably real</b>	<b>Analysed</b>
ANKA	ANKA	False positive (errors)	Filtered
	820	Line difference	Filtered
	234	Line difference	Filtered
	GNP (233, m7, m8 & m9)	<b>Probably real</b>	<b>Analysed</b>

**Table 4-2. A general schematic of how the total identified mutations were treated and classified for further analyses.** Basically, possibly erroneous/suspicious and lineage specific mutations were eliminated from the mutation pool by looking at the number of mutations in the GNPs which are common to their respective isogenic parental lines. The rest of the mutations were carried over for analyses considering them to be probably real. Mutations were initially NOT called in: regions flanking assembly gaps & mis-assemblies (telomeres), regions of low depth of coverage, non-coding regions except UTRs.

Thus, more than half (54%) of the SNPs were eliminated this way leaving behind 2004 potential SNPs and 184 indels (Table 4-3). These were further filtered for those co-occurring with mutations in lines 820cl1m1cl1cl2, 234 and/or ANKA as these indicate potential sequencing and/or assembly errors (false positive mutations) rather than real potentially causative SNPs and indels (FILTER 2).

The total number of mutations remaining after these 2 filters comprised 40 SNPs and 31 indels. A total of 6 synonymous SNPs were further filtered (FILTER 3) thus leaving 34 SNPs and 31 indels for further analyses. To this number were added 58 already filtered SNPs from the analysis of the GNP line 8, making the total number of SNPs 92. It is worth noting at this point that sequence comparisons between GNP lines and their isogenic cloned parent reveal a relatively small number of potential mutations. Considering the stability of these genomes, it is highly likely that these mutations have some potential consequences. The distribution of these across various lines studied (except K173) is shown in Table 4-3.

Table 4-3

COMPARISON AGAINST 820 REFERENCE													
LINES	TOTALS		FILTER 1 – Artefacts / errors		Eligible after F1		FILTER 2 – Common with 234, ANKA AND/OR 820		Eligible after F2		FILT 3 - Syn. SNP	Eligible after F3	
	SNP	IND	SNP	IND	SNP	IND	SNP	IND	SNP	IND		SNP	IND
233	1118	25	590	0	528	25	502	19	26	6	3	23	6
234	540	23	249	0	291	23	291	23	0	0	0	0	0
820	1008	40	521	0	487	40	487	40	0	0	0	0	0
ANKA	595	32	313	0	282	32	282	32	0	0	0	0	0
7	430	38	233	0	197	38	196	21	1	17	0	1	17
8	--	--	--	--	--	--	--	--	--	--	--	58	0
9	661	26	442	0	219	26	206	18	13	8	3	10	8
Total	4352	184	2348	0	2004	184	1964	153	40	31	6	92	31

**Table 4-3. Table showing total number of mutations within the coding sequences as identified by mapping lines 233, 234, 820cl1m1cl1cl2, ANKA, and lines 7, 8 and 9 on 820 reference and filters used to refine them. SNPs for line 8 were added as already filtered from the variant calling program and no indels were reported.**

### 4.3.2 *Mutations identified within the open reading frames in the GNP study lines after mapping onto the ANKA reference assembly*

In the same way as done for the mapping onto the 820 reference assembly, 1885 SNPs and 270 indels were identified in the predicted open reading frames (ORFs) when the lines were mapped onto the ANKA reference genome (Table 4-4). After eliminating the mutations which were also present (common mutations) in lines 820cl1m1cl1cl2 and/or 234 (FILTER 1), only 61 potential SNPs and 34 indels remained. A final filter to eliminate all synonymous SNPs (FILTER 2) refined the total number of SNPs and indels to 55 and 34, respectively. To this number were added 132 already filtered SNPs from the analysis of the GNP line 8, making the total number of SNPs 187. It is important here to note that after filtering for potential errors, only a small number of mutations (9 and 12, respectively) were finally observed in the *de novo* GNP lines 7 and 9, except in GNP line 8 where it totalled to 132 SNPs. Given the stability of the genomes, these mutations are likely to be significantly related to the observed phenotypic difference under study. The distribution of these remaining mutations across the studied lines (except K173) is shown in Table 4-4.

Table 4-4

COMPARISON AGAINST ANKA REFERENCE									
LINES	TOTALS		FILTER 1 – Common with 234 AND/OR 820		Eligible after F1		FILTER 2 – All synonymous SNPs	Eligible after F2	
	SNP	IND	SNP	IND	SNP	IND		SNP	IND
233	475	38	440	34	35	4	6	29	4
234	472	50	467	50	5	0	0	5	0
820	337	57	337	57	0	0	0	0	0
7	333	66	324	50	9	16	0	9	16
8	--	--	--	--	--	--	--	132	0
9	268	59	256	45	12	14	0	12	14
Totals	1885	270	1824	236	61	34	6	187	34

**Table 4-4: Total number of mutations within the coding sequences as identified by mapping lines 233, 234, 820cl1m1cl1cl2, and lines 7 and 9 on ANKA reference and filters used to refine them. SNPs for line 8 were added as already filtered from the variant calling program and no indels were reported.**

### 4.3.3 *Mapping mutations within 1 kb upstream and downstream of the open reading frames in the study lines using the ANKA reference assembly*

When the lines 233, 234, 820cl1m1cl1cl2, m7 and m9 were mapped onto the ANKA reference and mutations were called within 1 kb upstream and downstream of the coding sequence of all annotated genes (including the coding sequence of the gene), a total of 4533 SNPs and 1006 indels were identified (analysis not done for GNP line 8 because the GNP line 8 was only mapped on to line 820 and the data from WTSI was obtained by using a different version of SAM tools that included filtered mutations). Filtering them for those which were also present (common mutations) in lines 820cl1m1cl1cl2 and/or 234, left 107 and 147 potential SNPs and indels, respectively that are unique to the GNP lines (FILTER 1). A further filtering for synonymous SNPs (FILTER 2; as they do not result in an amino acid change) eliminated only 9 SNPs thus leaving 98 SNPs and 147 indels as potential candidates for further analysis although the final SNP list included 5 synonymous SNPs anticipated being important for further analysis. Their distribution across the studied lines (except K173) is shown in Table 4-5.

Table 4-5

COMPARISONS OF THE UTRs AGAINST ANKA REFERENCE									
LINES	TOTALS		FILTER 1: Common with 234 and/or 820		Eligible after F1		FILTER 2: Only Synonymous SNPs	Eligible after F2	
	SNP	IND	SNP	IND	SNP	IND	SNP	SNP	IND
<b>233</b>	1198	175	1141	149	62	26	5	57	26
<b>234</b>	1012	207	1009	202	3	5	0	3	5
<b>820</b>	827	199	827	197	0	2	0	0	2
<b>7</b>	739	199	724	150	17	49	2	15	49
<b>8</b>	NA	NA	NA	NA	NA	NA	NA	NA	NA
<b>9</b>	757	226	734	161	25	65	2	23	65
<b>Totals</b>	<b>4533</b>	<b>1006</b>	<b>4435</b>	<b>859</b>	<b>107</b>	<b>147</b>	<b>9</b>	<b>98</b>	<b>147</b>
Note: The SNPs eligible after F2 include 5 important synonymous SNPS									

**Table 4-5. Total number of mutations within the coding sequences & 1kb upstream and downstream in each of the coding sequences** as identified by mapping lines 233, 234, 820cl1m1cl1cl2, and GNP lines 7 and 9 on ANKA reference and filters used to refine them. Analysis was not performed for GNP line 8 because the GNP line 8 was only mapped on to line 820 and the data from WTSI was obtained by using a different version of SAM tools that included filtered mutations.



### 4.3.4 Summary of filtered mutations

Table 4-6 includes all the filtered mutations from tables 4-3 to 4-5. A grand total of 377 SNPs and 212 indels were carried forward for further detailed analysis. As Table 4-5 also includes mutations within the coding sequences (together with their 1 kb upstream and downstream UTRs), the common mutations between list 3 and lists 1 and 2 were thus eliminated (FILTER 1; but reserved for in depth analysis) leaving behind a total of 348 SNPs and 211 indels as shown in Table 4-7. A total of 277 SNPs and 184 indels were temporarily eliminated (FILTER 2) as being not known to be directly involved in gametocytogenesis and are thus pooled as reduced ranking priority mutations. These reduced ranking priority mutations mostly included those affecting conserved *Plasmodium* proteins with unknown functions (42%) and BIR genes (14%) and other miscellaneous gene products (Figure 4-2). Full details of these and their co-occurrence in the GNP lines are shown in Table A-11 in the Appendices.

Table 4-6

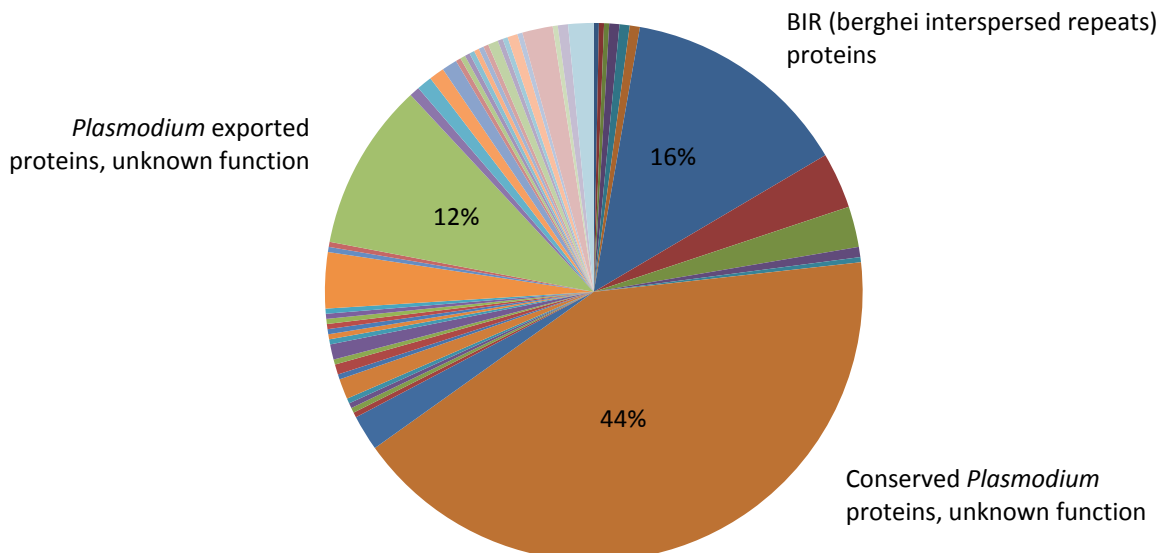
TOTAL FROM 3 LISTS								
LINES	TOTAL SNPs			TOTAL INDELS			GRAND TOTAL	
	LIST 1 - ON 820	LIST 2 - ON ANKA	LIST 3 - WITH UTRs ON ANKA	LIST 1 - ON 820	LIST 2 - ON ANKA	LIST 3 - WITH UTRs ON ANKA	SNPs	INDELS
233	23	29	57	6	4	26	109	36
234	0	5	3	0	0	5	8	5
820	0	0	0	0	0	2	0	2
ANKA	0	0	0	0	0	0	0	0
7	1	9	15	17	16	49	25	82
8	58	132	NA	NA	NA	NA	190	NA
9	10	12	23	8	14	65	45	87
<b>Totals</b>	<b>92</b>	<b>187</b>	<b>98</b>	<b>31</b>	<b>34</b>	<b>147</b>	<b>377</b>	<b>212</b>

Table 4-6. A summary of filtered mutations from lists 1, 2 and 3; tables 4-2, 4-3 and 4-5, respectively.

Table 4-7

FINAL MERGED LIST																
LINES	TOTALS		FILTER 1 - Common in UTRs & cds		Eligible after F1		FILTER 2 – reduced ranking priority mutations		Eligible after F2		FINAL PRIORITISED MUTATIONS					
	SNPs	INDELS	SNPs	INDELS	SNPs	INDELS	SNPs	INDELS	SNPs	INDELS	SNP			INDEL		
											IN CDS	IN UTR	TOTAL	IN CDS	IN UTR	TOTAL
233	109	36	9	2	100	34	92	28	8	6	7	1	8	3	3	6
234	8	5	2	1	6	4	1	2	5	2	4	1	5	0	2	2
820	0	2	0	0	0	2	0	1	0	1	0	0	0	0	1	1
7	25	82	8	0	17	82	17	72	0	10	0	0	0	7	3	10
8	190	--	--	--	190	--	134	--	56	--	56	--	56	--	--	--
9	45	87	10	0	35	89*	33	81	2	8	2	0	2	2	6	8
Tot	377	212	29	3	348	211	277	184	71	27	69	2	71	12	15	27
PS	The total number of SNPs includes 4 synonymous SNP and 1 SNP in the inter-exonic region															
	*ADDITIONAL MUTATIONS DISCOVERED															

Table 4-7. Final merged list of the number of mutations after filtering for reduced ranking priority mutations



**Figure 4-2. Breakdown of mutated gene products temporarily considered as reduced ranking priority and hence not included in the analysis.** The majority (44%) of these genes are conserved *Plasmodium* proteins (unknown functions), followed by BIR proteins (16%) and *Plasmodium* exported proteins (unknown functions; 12%).

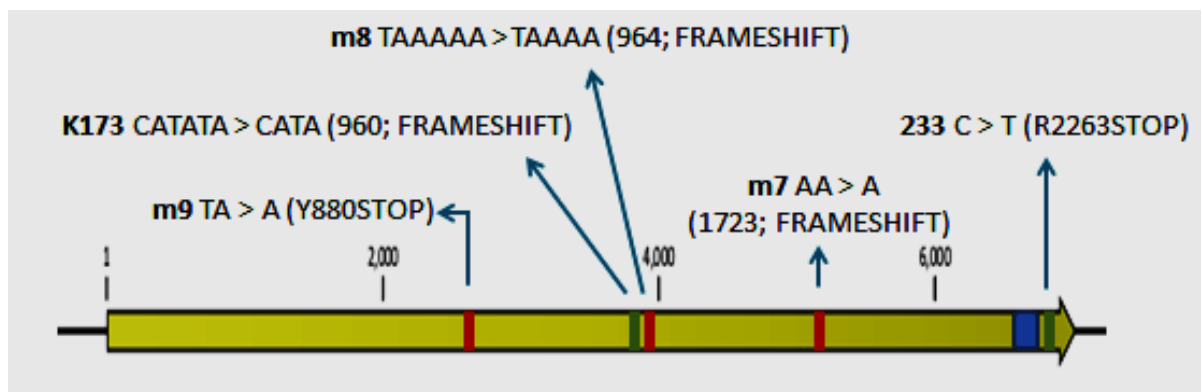
The final merged list of number of refined mutations included a total of 71 potential SNPs (69 in coding sequences and 2 in the UTRs) and 27 potential indels (12 in coding sequences and 15 in the UTRs). A breakdown of these across the studied lines is shown in Table A-12 in Appendices. These 71 SNPs and 27 indels were further characterized in detail and were grouped according to their co-occurrence (whether these mutations were present in isolated GNP lines or similar mutations in other GNP lines included in the present study) across other studied GNP lines as shown in Table A-12 in Appendices.

## 4.4 One gene uniquely mutated in all the GNPs

Out of the total number of genes (and their UTRs) found mutated, it is clearly evident that there is a single gene (PBANKA\_143750; a transcription factor with an AP2 domain) that has non-synonymous SNPs and/or indels resulting in premature insertion of a stop codon or transcriptional frame shift, respectively (Figures 4-1 and 4-3) in all GNP lines included in this study compared with the isogenic gametocyte producer parental genomes. The only other gene (PBANKA\_093910; another transcription factor with AP2 domain) that was found to be related to all the GNP under study had upstream alterations therefore PBANKA\_143750 was the overwhelmingly favourite candidate. However, there were other interesting candidates including RNA binding proteins, an RNA helicase, cdc2 related protein kinases, a histone demethylase, a translation initiation factor and a histone deacetylase among others but these occurred in isolation in one or the other gametocyte non-producer lines. Interestingly, none of these genes were found in common with the list of 15 genes that were found to be putatively responsible for gametocytogenesis in *P. falciparum* using transposon mutagenesis (Ikadai et al., 2013) except PBANKA\_040300 (ortholog of PFC0200w; 60S ribosomal protein; mutated in GNP m7) and PBANKA\_082360 (ortholog of PFI1215w; splicing factor 3a; mutated in GNP m7 and m9) - both the genes having mutations in 5' UTR of the ORF.

Whilst PBANKA\_143750 (hereinafter also referred to as *PbAP2-G* or *AP2-G*) stood out as the number one priority gene implicated in gametocytogenesis,

later confirmed by targeted gene deletion and successful gene complementation, others are certainly worth consideration in future as it can be hypothesised that it could be a cascade of genes/transcription factors rather than a single key determinant for commitment to gametocytogenesis. Also, considering the important role of epigenetics in non-transcriptional regulation of commitment (Brancucci et al., 2014; Coleman et al., 2014) together with a more contextual background in which the process of commitment occurs (Flueck and Baker, 2014), the implication and thus the importance of additional commonly associated genes cannot be completely ruled out.



**Figure 4-3. A schematic of the *AP2-G* (PBANKA\_143750) open reading frame.** The orf (yellow) is showing the location of the three naturally acquired *de novo* GNP mutations (m9, m8 and m7; maroon coloured vertical bars) and two pre-existing GNP mutations (K173cl1 and 233; green vertical bars) relative to the translation start site (position 1) of the *AP2-G* gene and the DNA binding domain (blue rectangle). These mutations were identified by mapping the reads from whole genome sequencing of the *de novo* and pre-existing GNP lines on the 820 reference assembly (see text). The details of the independent mutations are also shown as “reference base/s – altered base/s” pattern followed by the expected outcome of the mutation.

PBANKA\_093910 is another Api-AP2 family member gene, the 5'UTR of which is mutated in all the GNP lines characterised. The gene had a cluster of 4 different indels at various positions within 1 kb of the 5' UTR of the gene in all the gametocyte non-producer lines studied and compared to ANKA. Indels were also detected in parental lines 234 and 820cl1m1cl1cl2 when mapped on the ANKA reference assembly, however the indels differed between the GNP lines and their isogenic parental WT lines both in terms of their location and the bases involved (except in the 234/233 lines). When the gene and its UTRs were examined for any mutations in the available data from line K173cl1, 2 indels were detected within 1 kb 5' UTR. Mutations exclusive to the GNP involved the 2 indels in K173cl1 and one in line m9 (at -442 relative to the AUG). The rest of

the 3 indels, one each in line 233, m7 and m9 at position 1446814 (-132 relative to AUG), are same as the indel in line 234 whereas line 820 has an indel 1 base upstream at position 1446813. The functional significance, if any, of these is yet to be explored.

The mutations detected at the PBANKA\_093910 locus in K173cl1 are strikingly at different locations as compared to the mutations in other GNP lines and this could be explained on the basis of the lineage difference between K173 and ANKA. For the rest of the GNP lines, one possible explanation of how mutations upstream of the ApiAP2 proteins might regulate gene expression is by interrupting the sequence of a DNA binding motif so that the DNA binding domain is prevented from binding to its cognate DNA sequence. Although, no motif consensus sequence (GTAC) was identified around the position/s where indels were reported, the identification of the target sequence GTAC at position -300 relative to AUG, poses intriguing possibilities that could attribute the indel at position -442 relative to AUG in GNP line m9 to the GNP phenotype in that line possibly due to altered ability of PBANKA\_143750 DBD to bind here. It has been shown that the *P. falciparum* homologue to PBANKA\_093910 (PF11\_0091) recognizes a specific motif (aGcATAC) to bind DNA (Painter et al., 2011). A simple motif search with GCATAC found no such sequence within 1 kb upstream to the AUG suggesting that this ApiAP2 might not influence its own expression and hence the indels detected above might not affect transcription in an autoregulatory manner.

#### 4.4.1 Evidence from pre-existing GNP lines

Apart from the identification of a single gene (PBANKA\_143750) uniquely mutated in all the GNP lines under study, whole genome resequencing of the pre-existing GNP lines, 233 and K173, identified another two ApiAP2 transcription factors (PBANKA\_011210 and PBANKA\_141570) mutated in both pre-existing GNP lines.

#### **4.4.1.1 Mutations in PBANKA\_011210**

Whole genome resequencing of the two pre-existing GNP lines, 233 and K173, that were included in the study revealed that another ApiAP2 transcription factor (PBANKA\_011210) was mutated in both of these lines. The translated protein product of PBANKA\_011210 has a DNA binding domain ranging from 1813 to 1866 amino acids of the 2578 aa protein. The GNP line 233 has a thymine to cytosine non-synonymous SNP at position 453650 bp resulting in an amino acid change from phenylalanine to serine at position 1823. The gene, when examined for mutation in line K173cl1 revealed the same mutation. The same mutation was however also detected in the 234, the gametocyte producer parental line for 233. It is important to note, however, that the mutation is present within the DNA binding domain of the protein and hence might be a candidate for further exploration.

#### **4.4.1.2 Mutations in PBANKA\_141570**

PBANKA\_141570 was yet another transcription factor with AP2 domains with mutations detected in both the pre-existing GNP lines - 233 and K173cl1 together with mutation in the parental line 234. The gene had a G to A SNP at position 591556 bp in the two GNP lines, along with the same mutation in line 234. The mutation was just upstream to the stop codon for the gene and involved the last amino acid, changing it from methionine to isoleucine. The DNA binding domain of the protein spans amino acids 1156 to 1209 of the 2487 aa long protein and was thus 3834 bp upstream to the locus of mutation. The functional interpretation of these mutations is similar, in principle, to that for the gene PBANKA\_011210.

### **4.5 Confirmation of PBANKA\_143750 mutations by direct sequencing**

The mutations identified by whole-genome re-sequencing were confirmed by di-deoxy sequencing to eliminate any association with sequencing errors. The full open reading frame of PBANKA\_143750 (7.02 kb) was amplified from the

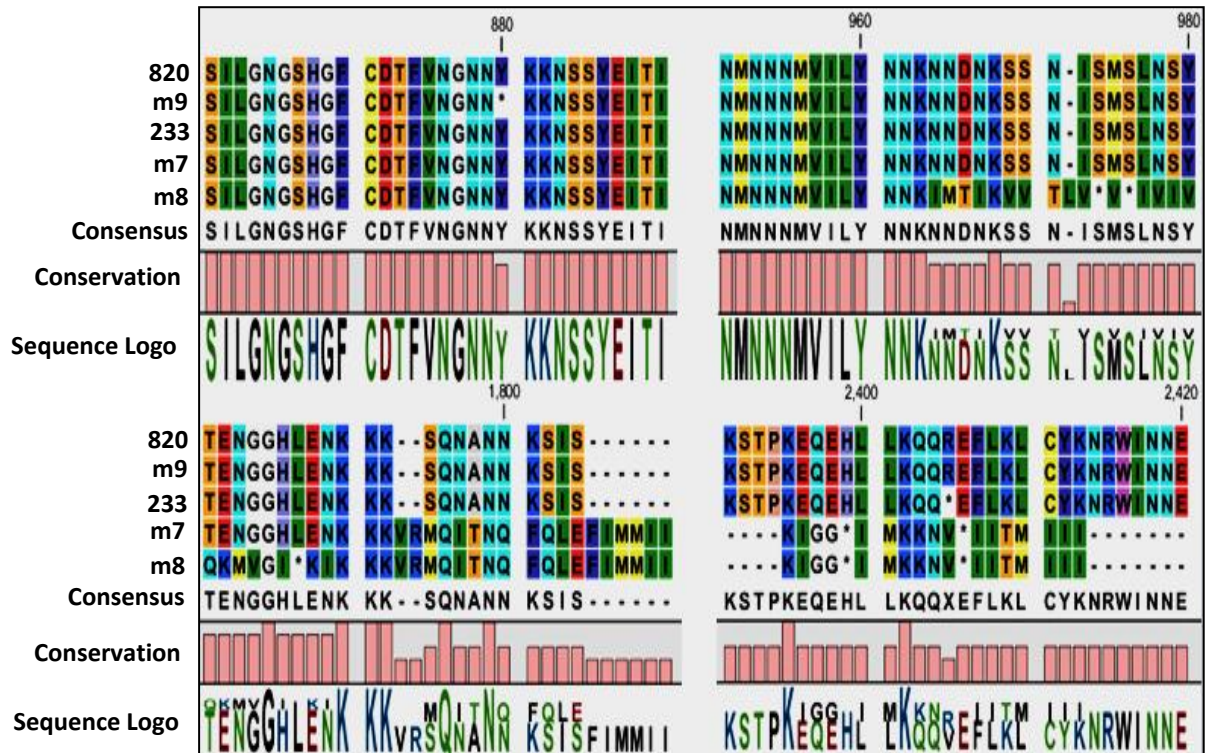
purified genomic DNA of WT line 820, the three *de novo* GNP lines (m9, m8 and m7) and the pre-existing GNP lines (233 and K173cl1) using primers GU1070/GU1071 and a proof-reading polymerase (HiFi Long template Taq Polymerase; Roche) and sent for sequencing to Eurofins MWG operon, Germany. In case, if the full orf was not amplifiable, the gene was amplified in 4 different parts of ~2 kb each and all the fragments sequenced. The following primer pairs were used for fragment 1 (GU1070/GU1239), fragment 2 (GU1182/GU1240), fragment 3 (GU1241/GU1242) and fragment 4 (GU1237/GU1071).

The DNA sequences were first converted to protein sequences and were subsequently aligned using CLC Genomics Workbench version 5.1 software (Figure 4-4). The sequence alignment confirmed the mutations in the various GNP lines studied and also ruled out the presence of any mutation/s other than those detected by the WGS and subsequent mapping.

## 4.6 The ApiAP2 family of transcription factors and AP2-G

Interest in transcription factor-based gene regulation in *Plasmodium* has arisen from a recent discovery of a family of conserved proteins containing *apicomplexan* AP2 DNA-binding domains (known as the ApiAP2) (Balaji et al, 2005; Yuda et al, 2009; Yuda et al, 2010). Three of the conserved *P. berghei* AP2 genes have already been implicated in controlling stage specific development in *Plasmodium* (Yuda et al., 2009; Yuda et al, 2010; Iwanaga et al., 2012). Further attempts are continuing to pinpoint the functional roles of various ApiAP2 proteins in transcriptional regulation during *Plasmodium* development and as these are strong candidates as potential regulators of gene expression in *Plasmodium*.

PBANKA\_143750 is a 7020 bp long gene and spans from genomic position 1365549-1372568 bases on the forward strand on chromosome 14 of the most recent version of the *P. berghei* genome assembly ([ftp://ftp.sanger.ac.uk/pub/pathogens/P\\_berghei/February\\_2011/](ftp://ftp.sanger.ac.uk/pub/pathogens/P_berghei/February_2011/)) and contains 2339 amino acids. The gene has a 55aa DNA binding domain (characteristic of the family of ApiAP2 transcription factors) spanning from 2197-2252aa.

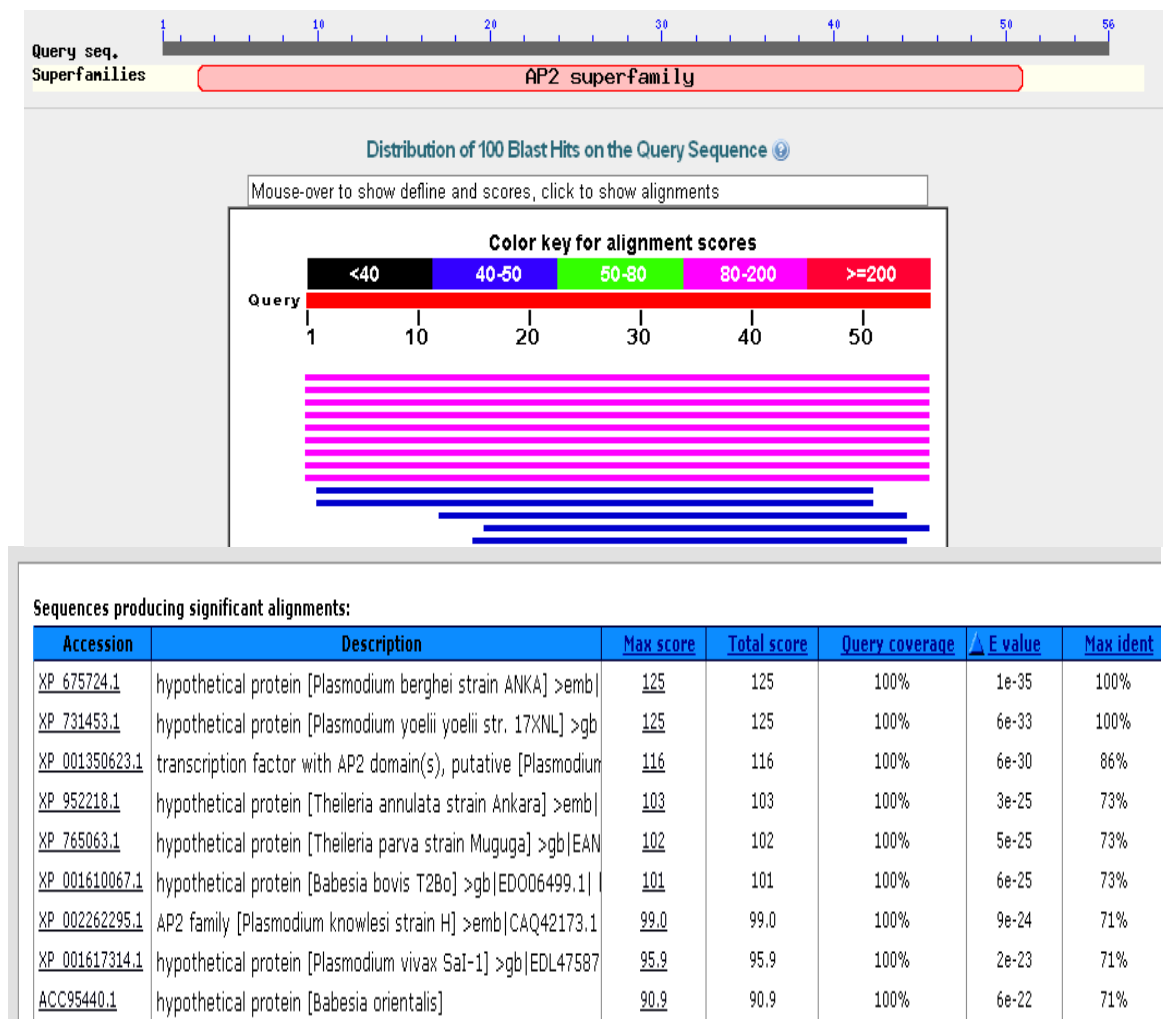


**Figure 4-4. Portions of multiple amino-acid sequence alignment (CLC Genomic Workbench software version 5.1) converted from nucleotide sequences obtained from sequencing the PCR amplified copy of the WT full length PBANKA\_143750 from the parental line 820 and GNP lines m9, m8, m7 and 233.** The consensus sequence, conservation of the amino-acids and the sequence logo are shown below the alignment. Only the parts of the alignment which had the mutations in the GNP lines are shown but there were no mutations in the gene other than that depicted here corroborating the findings of the whole genome resequencing approach. The line m9 had a stop codon (\*) at position 880 whereas 233 had the same at position 2263. Lines m7 and m8 had frameshifts at positions 1723 and 974, respectively.

#### 4.6.1 AP2-G DBDs are highly conserved across apicomplexa

The DNA binding domain (DBD) and target recognition motif of the *PbAP2-G* are highly conserved across orthologues in even distantly related apicomplexans. A BLAST search for the AP2-G DBD query sequence resulted in a 100% identity (top 2 hits) to the two rodent malaria parasites, *P. berghei* and *P. yoelii* (Figure 4-5). The next four similarity hits in decreasing order of identity were *P. falciparum* (86%), *Theileria annulata* (73%), *Theileria parva* (73%) and *Babesia bovis* (73%). This conservation suggests that the ApiAP2-G TF play important and critical roles during the life cycle of apicomplexans.





**Figure 4-5. BLAST search hits when AP2-G DBD sequence of *P. berghei* was used as the query sequence.** The top two hits showed 100% identity as they belonged to the rodent malaria parasites – *P. berghei* and *P. yoelii*. The next four hits in decreasing order of similarity were *P. falciparum* (86%), *Theileria annulata* (73%), *Theileria parva* (73%) and *Babesia bovis* (73%).

Since PbAP2-G DBD sequence identity on BLAST is more with some of *Theileria* and *Babesia* sequences as compared to the identity with *P. knowlesi* and *P. vivax*, it appears that PbAP2-G DBD forms a phylogenetic cluster with orthologues from *Plasmodium*, *Theileria* and *Babesia* species and thus may be functionally associated with sexual differentiation in these species. The restriction of the orthologues to the ones with a clearly defined asexual stage implies that the AP2 transcription factor has a deterministic role during the sexual stages of these organisms.

It has been shown that *Plasmodium* (*P. falciparum*, *P. berghei* and *P. yoelii*) and *Theileria* share a higher number of orthologous AP2 domains between them within *apicomplexa* supporting conclusion that these are phylogenetically

relatively closer (Leander et al., 2003; Balaji et al., 2005). It might be possible that certain AP2 domains regulating a particular stage in the life cycle of certain apicomplexans evolved prior to the diversification of their genera (Brian Shiels, University of Glasgow, personal communication). Recent research with *Theileria* has demonstrated that an ApiAP2 protein containing the clear orthologue of the DBD from PbAP2-G in *Theileria annulata* (TA13515) is up-regulated in piroplasms prior to the production of gametes (Brian Shiels and Marta Pieszko, University of Glasgow, personal communication).

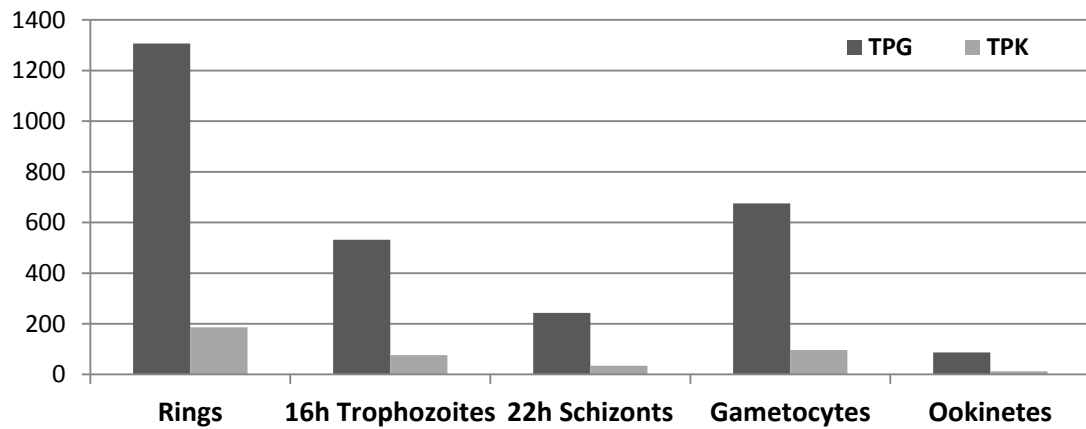
### 4.6.2 Expression profile of PbAP2-G

The expression profile of *PbAP2-G* through various developmental stages of *P. berghei* including ~4-hour rings, 16-hour trophozoites, 22-hour schizonts, gametocytes and ookinetes was analysed from the RNA sequencing database available with the Waters lab (Hoeijmakers WAM, Religa A, Janse CJ, Waters AP, and Stunnenberg HG, unpublished data) generated from purified mRNA from each of these stages obtained from synchronized infections.

Gene expression is shown either as the number of tags per gene (TPG) or number of tags per kilobase (TPK) of the genome. Although both measure the amount of gene expression, TPG shows a higher value because of the presence of intergenic areas which don't contribute to gene expression and hence dilute the TPK value (Figure 4-6). Although the number of tags per gene is the highest for the ring stage, comparative expression ranking of *PbAP2-G* amongst all the genes expressed within the ring stage is very low at 21 percentile which means that only 20% of the genes expressed at ring stage rank below *AP2-G* expression. Similarly, the expression ranking of *AP2-G* within the 22-hour schizont stage is relatively better - 67 percentile. The corresponding ranking at 16-hour trophozoites and gametocytes is 50 percentile and 30 percentile, respectively.

With such relatively low expression of *PbAP2-G* in all stages of the intra-erythrocytic developmental cycle of the parasite, it was difficult to assess the expression and localisation of the gene, not even with a tagged transgene (Sinha et al., 2014)). The expression profile of the *AP2-G* transcription factor appeared to be comparatively better at the late schizont stage. With an implicated role of

AP2-G in commitment to sexual stage differentiation in malaria parasites, this relatively higher expression of the gene would suggest that the commitment might take place during or just following the late schizont stage in *P. berghei*.



**Figure 4-6. Expression profiling of *PbAP2-G* across different developmental stages** of synchronised *P. berghei* parasite stages shown as the number of tags per gene (TPG) and tags per kilobase of the genome (TPK). From the data, it appears that *PbAP2-G* is most highly expressed in the ring stage followed by the gametocytes.

### 4.6.3 *Localization studies; GFP- and HA-tagged PbAP2-G*

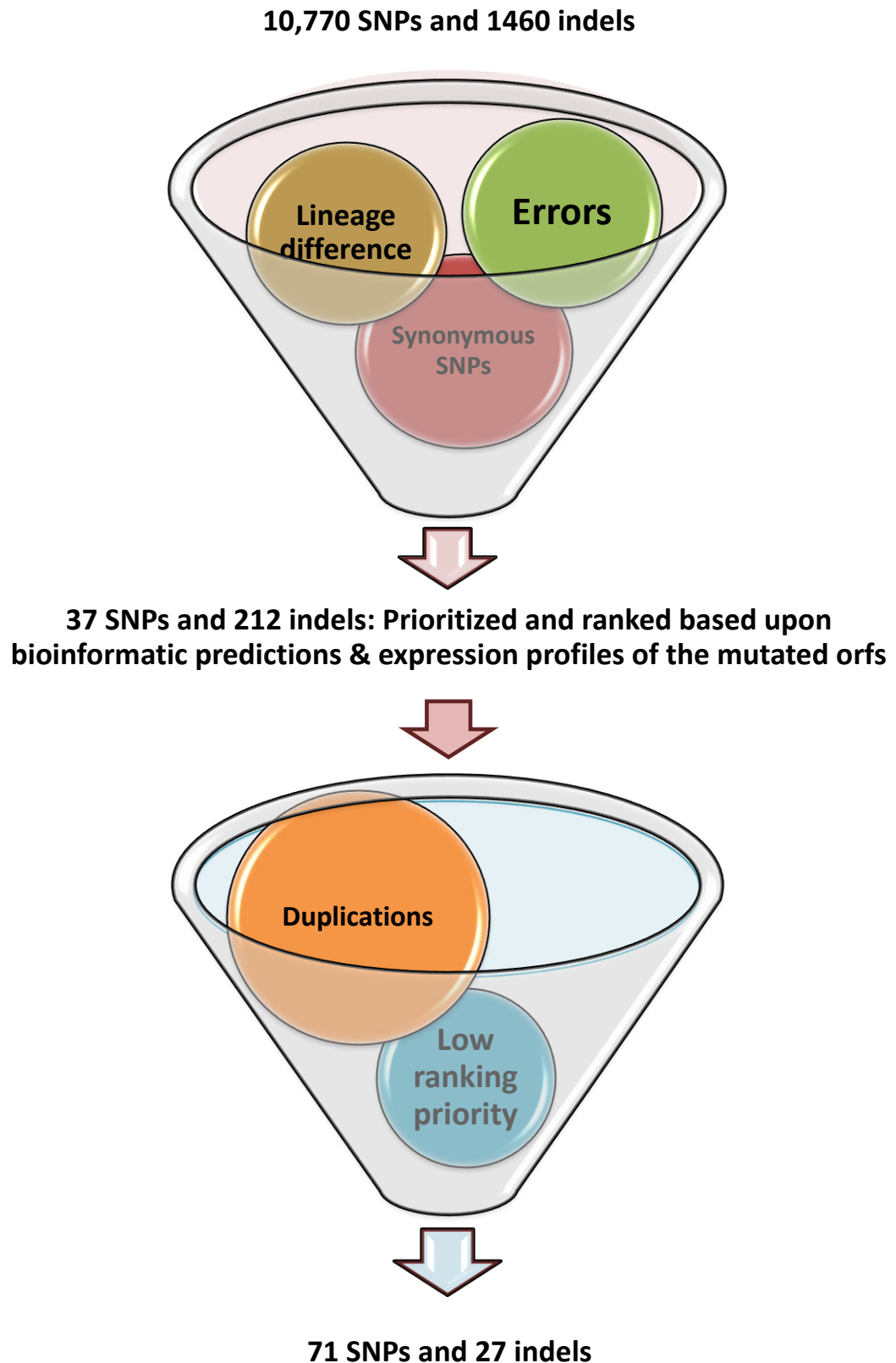
Although efforts to epitope tag AP2-G with GFP and HA were made, success could not be achieved in tagging the gene. One of the group members (Anne Graham; data not shown) did eventually succeed in tagging allelic *PbAP2-G* with both GFP and HA yet the protein could not be detected even with available peptide antisera or commercial antibodies against the tags, possibly because it is expressed at low levels. However, a truncated transgene was generated by yet another member of the group (Katie Hughes; data not shown) comprising the N-terminal 100aa and the entire C-terminal region (300 aa) including the DBD sandwiching CFP and driven by the *PbAP2-G* promoter (2 kb of upstream DNA). The transgene could be detected in nuclei of female gametocytes, consistent with a role for *PbAP2-G* in gametocyte development.

## 4.7 Summary and discussion

### 4.7.1 *Whole-genome sequencing and GNP mutations*

Whole-genome sequencing (WGS) based studies in *Plasmodium* have become increasingly popular in the last decade after the publication of the malaria genome in 2002 (Gardener et al., 2002; Bright and Winzeler, 2013). Although the majority of such studies emphasized the discovery of the molecular basis of drug resistance in *Plasmodium* (Miotto et al., 2013; Neafsey, 2013; Bright and Winzeler, 2013; Menegaon et al., 2007; and Eckland and Fiddock, 2007), these high throughput technologies could be harvested for better understanding of basic parasite biology and identification of novel drug and/or vaccine targets.

WGS was thus applied in the current study towards understanding the molecular basis of commitment to gametocytogenesis in malaria parasites. Three natural GNP mutants were generated by serially passaging a WT fluorescent labelled gametocyte producer clone and their genomes were sequenced. The sequences from these lines and the two existing GNP lines, 233 and K173, were compared with the WT parental line 820 and line ANKA. Mapping of the sequencing reads from the WGS of the three GNP lines on to their parental reference assembly identified approximately 10,770 SNPs plus 1460 indels. These were then sequentially filtered (Figure 4-7) to a final list of 15 SNPs and 27 indels. On the basis of commonality of occurrence in all the GNP lines, PBANKA\_143750 (transcription factor with AP2 domain) emerged as the one and the only gene to be uniquely mutated in all three *de novo* GNP mutants plus the pre-existing GNP lines K173cl1 and 233.



**Figure 4-7. A flow diagram showing the screens applied to filter the detected mutations.** The screens eliminated false positive mutations (being detected due to lineage differences between the lines mapped and variant calling errors), synonymous SNPs (resulting in no amino acid change), duplication in mutation detection, and temporarily, the mutations in genes classified as reduced ranking priority on the basis of bioinformatic prediction of gene function and possible association with gametocytogenesis.

### 4.7.2 ***De novo GNP mutations and PBANKA\_143750 (AP2-G)***

Apart from the evidence generated in this study which establishes a possible role of PBANKA\_143750 in commitment to gametocytogenesis, compelling evidence resulting from a parallel and independent approach using *P. falciparum* also identified the same ApiAP2 transcription factor (PFL\_1085w) as the critical regulator of commitment to gametocytogenesis in malaria parasites (Kafasack et al., 2014).

Crucial evidence in the present study comes from the isogenic *de novo* GNP lines m7, m8 and m9 derived from the parental gametocyte producer line 820cl1m1cl1cl2. GNP line m9 had a non-synonymous SNP at position 2640 causing the base change from thymine to adenine with the resulting amino acid change at position 880 from tyrosine to a stop codon. Whereas in the GNP lines m7 and m8, there are indels (AA to A) at position 5165 in m7 and (TAAAAA to TAAAA) at position 2888 (relative to the translation start site) which result in a transcriptional frame shift. In both of these situations, this is expected to produce an altered functional product - a truncated protein devoid of its DNA binding domain and thus probably ineffective as a transcriptional regulator. The predicted sizes of the truncated protein product of the naturally mutated PBANKA\_143750 was 879aa in m9, and 1759aa in m7 (see Table 4-8), as compared to 2339aa in the WT gene.

An important functional consideration while exploring these mutations is the position of the mutation on the gene relative to the DNA binding domain. With the exception of line 233, the mutations in the rest of the 3 GNP lines studied here occur upstream of the DNA binding domain. This relative location of the mutations has a functional implication - the nonsense/missense mutations occurring upstream to the DNA binding domain are expected to result in a truncated/altered protein product, respectively, without the DNA binding moiety. Such upstream mutations disrupting the function of AP2 domain would be manifest at the transcriptional level. In contrast, a nonsense mutation downstream to the DNA binding domain in a GNP (as in line 233 - R2263STOP) is expected to result in a fully functional DNA binding domain. This might indicate

yet another level of functional regulation which may involve protein-protein and/or protein-DNA interactions involving the region of the AP2 protein downstream of this mutation.

Table 4-8

Line	Nucleotide Position*	Type	Nucleotide change	Amino acid change	Effect	Predicted length of the protein (aa)#
7	5165	Indel	AA to A		Frameshift	1759
8	2888	Indel	TAAAAA to TAAAA		Frameshift	973
9	2640	SNP	TAT to TAA	Y to STOP	Termination codon	879
233	6787	SNP	CGA to TGA	R to STOP	Termination codon	2262
K173	2880	Indel	CATATA to CATA		Frameshift	959

**Table 4-8. Distribution of mutations across PBANKA\_143750 orf.** The table shows the distribution with respect to the nucleotide position and nucleotide change involved in the mutation, the type of mutation, the resulting amino acid change, the predicted effect as a result of the mutation and the predicted length of the mutated protein. \* With respect to ATG; # Predicted length of WT protein: 2339aa.

Recent evidence from protein binding microarray experiments involving recombinant AP2 fragments of PFL1085w, the PBANKA\_143750 orthologue in *P. falciparum* (Manuel Llinas, personal communication) suggest that extra sequences both at N- and C-terminal to the DNA binding domain (35aa upstream and 32aa downstream to the domain) in addition to the domain itself, are critical for DNA binding (Balaji et al., 2005; Campbell et al., 2010). In the context of the mutation in GNP line 233, it is important here to note that the mutated amino acid at position 2263 falls within this downstream critical zone (domain plus 32aa downstream) and thus might explain the GNP phenotype. Using such recombinant proteins as a guide, it can be speculated that a mutation in the AP2 protein downstream of its DNA binding domain might still affect its capacity to bind cognitive DNA sequence. Additional evidence from existing *P. berghei* GNP lines and/or their *P. falciparum* counterparts will certainly be crucial to explain the possible mechanisms involved in the expression of genes

likely to play a contributory role in commitment to gametocytogenesis in *Plasmodium*.

### **4.7.3      *Supplementary evidence from WGS of pre-existing GNP lines***

The involvement of PBANKA\_143750 as a critical regulator of commitment to gametocytogenesis in malaria parasites is further supported by analyses of the WGS data from the already available GNP line 233, wherein PBANKA\_143750 has a non-synonymous SNP at position 6787 (relative to the translation start site) which causes a single nucleotide change from cytosine to thymine thus changing the amino acid at position 2263 from arginine to a stop codon thus potentially precociously terminating the transcription. Supplementary evidence comes from the available data of an independent already available unrelated non-gametocyte producer line K173cl1, where the gene has an indel at position 963 (relative to the translation start site) thus generating a transcriptional frame shift (CATATA-CATA) and thus an altered product. K173cl1 had a high number of WGS SNPs and indels when mapped to ANKA, but this is expected as K173cl1 has a lineage which is different from ANKA.

One of the important considerations to be borne in mind while attributing differentially identified mutations in the pre-existing GNP lines to the GNP phenotype includes the fact that various mutations occur in the parasite genome over time due to continuous evolution of the parasite and there is thus a tendency of the parasite to accumulate the mutations that remain favourable to the parasite during their evolutionary development. Because the pre-existing GNP lines have an unknown number of generations apart from the isogenic parental clone, the precise time of evolution of the GNP phenotype remain unascertained, mutation/s may be spuriously associated with the phenotype and hence may not be real. Any phenotype associated with mutations in such a lineage should always be confirmed by reverse genetic experiments to causally link the ascribed mutation/s to the phenotype. Recent data, however, shows that the basal rate of accumulation of mutations in *P. falciparum* remains low ( $1.0\text{--}9.7 \times 10^{-9}$  mutations per base pair per generation) and also that such



mutation/s mostly remain confined to sub-telomeric multi-gene families with a relatively stable core genome (Bopp et al., 2013).

This has three implications - either the mutation has no role towards commitment to gametocytogenesis in these two lines (at least) or it could also be possible that, since 234 is also derived from ANKA, although the line 234 itself is a producer of gametocytes it is actually in a transitional phase in a continuum to develop gametocyte non-producer mutants from ANKA. If this might be the case, it will be intriguing to explore the point at which this mutation is acquired and the post-acquisition GNP kinetics of the line, especially when it is not known how many generations the lines 234 and 233 are apart. The third possibility could be that the process of commitment to gametocytogenesis is cascade dependent and GNP phenotype occurs through mutations in a series of genes involved, of which, PBANKA\_011210 is a part, the master regulator being PBANKA\_143750. This would then imply that PBANKA\_011210 is positioned upstream to PBANKA\_143750 *AP2-G* in the hypothesised cascade. This places more emphasis on the data derived from the *de novo* gametocyte non-producer mutants in the present study where we can explore the time point/s of acquisition of any particular mutation simply by retrospectively tracking/analysing the genome sequence of a particular GNP line. Efforts to knockout PBANKA\_011210 were attempted but a successful knockout could not be generated. The efforts could not be continued / repeated due to paucity of time.

Further, the Waters lab already possesses the following available pre-existing GNP lines HPEcy1m50cl1, PbHpD25mcl17, PbHpD25mcl20, Pb LPjaune8409, Pb LPvert8458 (kind gift from Chris Janse, Leiden University) and KSP11, SP11 and LUKA (kind gift from Sarah Reece, University of Edinburgh). The gene of interest, PBANKA\_143750 and other genomic areas of potential interest pertaining to gametocytogenesis regulation could be amplified for sequencing. Any mutations in *AP2-G* (PBANKA\_143750) and/or other examined genomic regions in comparison to the ANKA reference could provide conclusive supporting and contextual evidence towards the implication of PBANKA\_143750 and its precise functioning in commitment to gametocytogenesis.

**5 Gene knockout studies to confirm the involvement of AP2-G in commitment to gametocytogenesis**

## 5.1 Introduction

Genetic manipulation, in terms of gene knockout, gene knock-in and gene disruption is one of the essential and fundamental molecular biology tools to investigate the role of the implicated gene/s

As the gene under study was longer than the usual (~7 kb), a multitude of targeted gene knockout strategies were planned and carried out. These included conventional double- and single- crossover designs and also the newer recombineering-based knockout vector designs. Stable transfection of the *Plasmodium* schizont stage with genetic constructs that allow homologous recombination based targeted gene alterations are now standard procedures (van Dijk et al., 1995; Wu et al., 1996; Fidock et al., 2000; Crabb, 2002; Sidhu et al., 2002; Su et al., 2007).

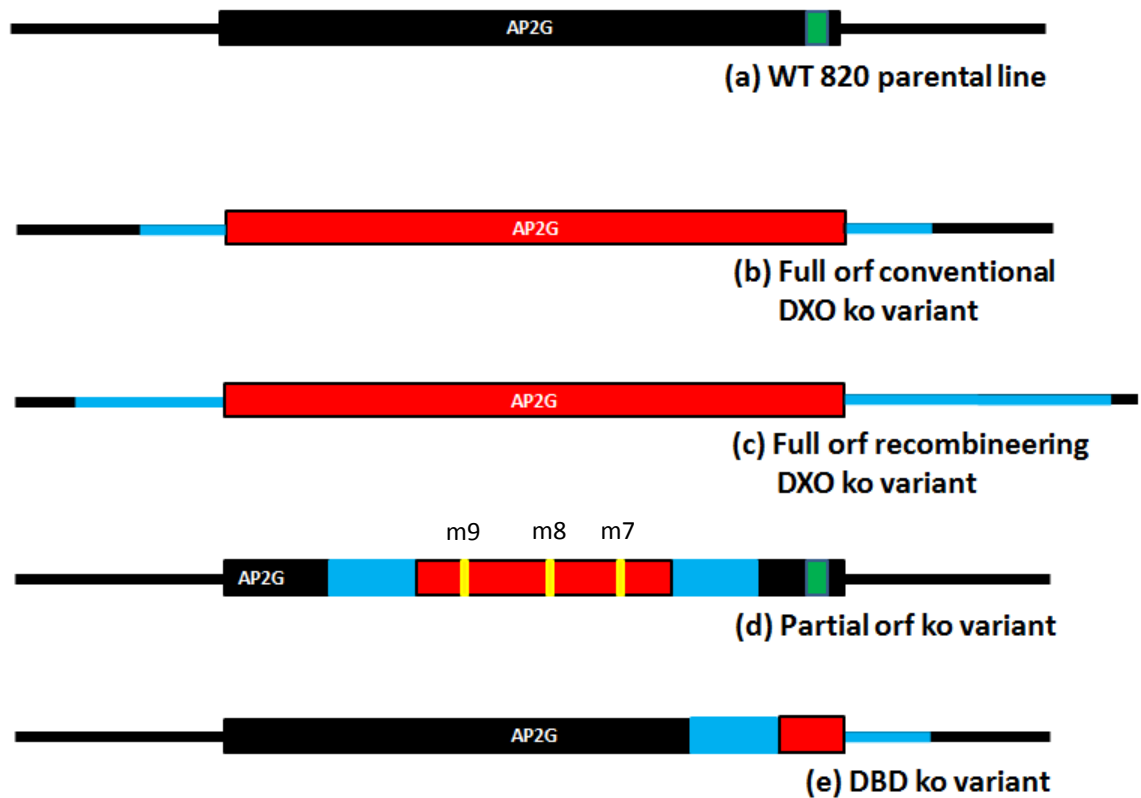
Gene knockout strategies in *Plasmodium* are based on the principles of homologous recombination between the targeted genomic sequence and its homolog present in the gene-targeting vector. Conventionally, this vector construct could, in turn, be circular (plasmids; as in *P. falciparum*) or a linearized piece of DNA (as in *P. berghei*) and carry a single copy of the homology arm upstream of the drug selectable marker (insertion type) or two copies of the homology arms, flanking a drug selectable marker (replacement type). The length of the homology arms is usually between 0.4-1.0 kb.

In *P. berghei* targeted genetic manipulation, the insertion-type vector is prepared for transfection by linearizing it with a unique restriction enzyme and the linearized DNA construct inserts into the genome by single crossover (SCO) between homologous arms - the vector backbone being retained in the modified genome. In contrast, the replacement-type vector is prepared by linearizing the vector with two restriction enzymes which removes the intervening vector backbone and the linearized DNA construct inserts into the genome via a double crossover event (DCO) between the two homologous homology arms thereby replacing the target region with the drug selectable marker (Carvalho and Menard, 2004; Janse et al., 2006).

Alternatively, the DCO gene targeting vector could be a linear construct instead of a circular plasmid that is derived from bacteria (bacteriophage N15 which can be efficiently modified using the lambda Red method of recombineering) and containing very long homology arms - up to 10 kb (Pfander et al., 2011). Since these newer recombineering-based approaches claim to improve the recombination frequency substantially (Hasty et al., 1991; Shulman et al., 1990), particularly in highly AT-rich genomes like *Plasmodium* (Pfander et al., 2011), these were included as strategies for gene knockout experiments with *Pb AP2-G*.

## 5.2 The *AP2-G* knockout length variants

As discussed in the previous section regarding the potential expected difficulties in knocking the whole open reading frame (ORF, 7.02 kb) out, the conventional double-crossover targeted knockout was attempted for various lengths of the ORF of *AP2-G*. The full length knockout of *AP2-G* involved the attempted knockout of the whole 7020 bp of the ORF with the help of ~1 kb upstream and downstream conventional short homology arms (all homology arms are indicated in blue in Figure 5-1) and also with longer homology arms (~1.8 kb and 2.7 kb; the recombineering-based approach) all situated respectively in the 5' and 3' UTRs of the gene (Figures 5-1b and c). The partial gene knockout (also named part-knockout) involved the knockout of a region of the orf bearing the m7, m8 and m9 mutations (these mutations were situated between 2167 bp to 5636 bp from the ATG; a total of 3470 bp of the ORF). The homology arms contributing to the partial gene knockout were derived from the regions of the *AP2-G* ORF flanking the mutations (Figure 5-1d). A third partial gene knockout was designed to delete the DNA binding domain (DBD) involving the full length DBD of *AP2-G* (from 6066 bp to 7020 bp from the ATG; total 954 bp including 168 bp DBD) with the upstream homology arm coming from within the orf flanking the DBD and the downstream homology arm derived from the 3'UTR of the gene (Figure 5-1e).



**Figure 5-1 (a-e). Schematic representation (not-to-scale) of the various length variants of AP2-G knockouts in *Plasmodium berghei*.** The genomic location of the *PbAP2-G* (PBANKA\_143750) is represented in (a). The black thicker bar represents the open reading frame (orf) of AP2-G and the green rectangle of the orf represents the DNA-binding domain (DBD). The different lengths of the AP2-G orf knocked out are shown in red bars placed upon various portions of the AP2-G orf including full-length orf ko (b and c), partial length orf ko (d) and DBD ko (e). Blue bars (b-e) represent different lengths of the homology arms used for homologous recombination: thin blue bars represent homology regions outside the AP2-G orf whereas blue thick bars represent homology regions within the orf. The relative length of the thin blue bars in (b) and (c) highlight the importance of recombineering based ko approach with 1.8 kb and 2.7 kb long homology arms in (c) as compared to conventional ~1 kb long homology arms in (b), (d) and (e). Yellow vertical bars in (d) represent the relative location of the three gametocyte non-producer (GNP) mutations which developed during the weekly passage of the WT line.

## 5.3 The knockout genotype

The success of the knockout transfectants attempted was initially assessed with the diagnostic PCRs and finally confirmed by PCRs and Southern analyses of the cloned mutants.

### 5.3.1 *PCR analyses of the knockout mutants:*

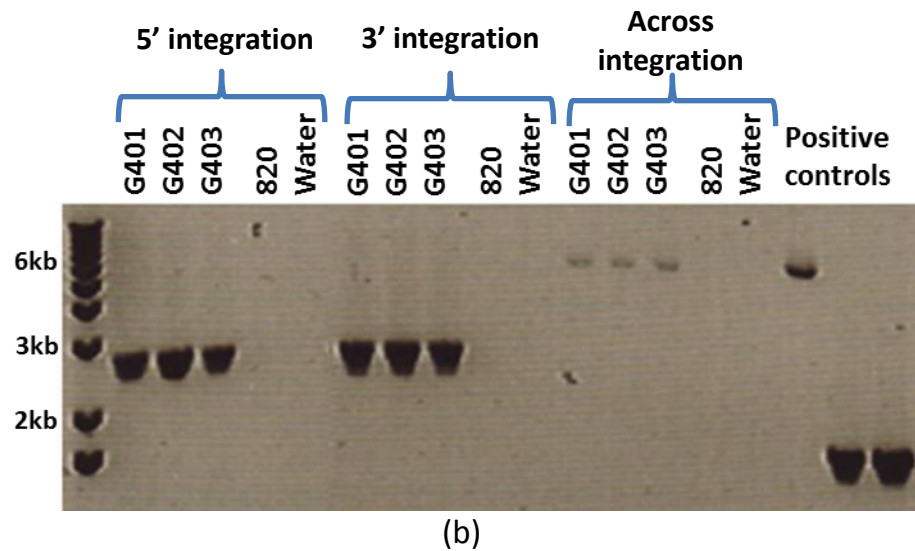
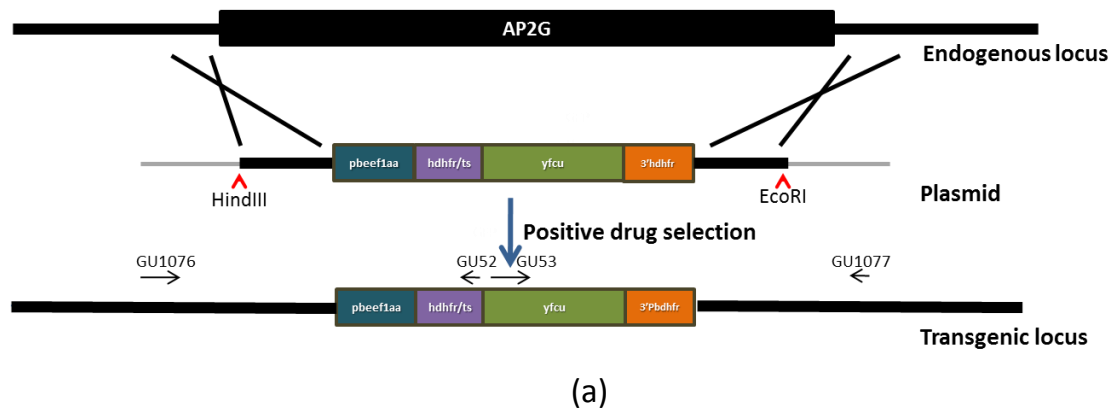
#### 5.3.1.1 The double-crossover (conventional) full length *AP2-G* orf knockout mutant

HindIII and EcoRI double-digested (New England BioLabs), linearized and purified targeted vector DNA (pG0271) was transfected (~5 µg) to the purified schizonts of the WT line (820em1dcl2) in triplicate. The mutant lines (G401, G402 and G403) were selected by using pyrimethamine in drinking water. Diagnostic PCRs were performed using oligonucleotide pairs GU1076/GU52 (for 5' integration) and GU53/GU1077 (for 3' integration) as shown in Figure 5-2. The expected approximate length of the amplicons for each of the integration events was 2.6 kb (Table 5-1). The transfectants were also screened by amplifying for the full length of the region knocked out by using oligonucleotide pairs GU1076/GU1077 (expected band size 5.4 kb in the ko mutants and 9.0 kb in the WT) and line G401 was taken further for cloning. A summary of PCR details is shown in Table 5-1. Presence of integration bands of expected size in all three transfected lines suggested a successful *AP2-G* knockout.

Table 5-1

Oligo set	Details	Expected band size with DNA (kb)				Ta (°C)	Ext. Time (min)
		820	G401-3	Water	HP		
1076/52	5'integration	--	2.6	--	NA	50	5
53/1077	3'integration	--	2.6	--	NA	54	5
1076/1077	Across integration	9	5.4	--	NA	54	10
1233/1235	Control1	NA	NA	NA	5.9	54	10
1233/1242	Control2	NA	NA	NA	2.9	50/54	5

**Table 5-1. Basic PCR reaction settings for amplifying various regions of the *PbAP2-G* genomic locus to identify the presence of 5' and 3' integration of the knockout constructs into the genome. Ta: annealing temperature for the primer pair**



**Figure 5-2 (a).** Schematic of *AP2-G* (PBANKA\_143750) conventional DXO knockout strategy (not-to-scale) showing endogenous & modified genetic locus together with the plasmid which was linearized with *HindIII* & *EcoRI* before transfection in triplicate into the WT 820 line. The resulting transgenic mutants (G401-3) were positively selected using the drug pyrimethamine. Also shown are the positions of the oligonucleotide primers (GU1076/GU52) and (GU53/GU1077) used for examining the successful integration of the construct into the genomic locus. **(b)** Diagnostic PCR confirmed successful integration using the following primer pairs: (1; 5' integration=2.6 kb) GU1076/GU52 (2; 3' integration=2.6kb) GU53/GU1077 and (3; through the modified locus=5.4 kb in the ko and 9.0 kb in the WT) GU1076/GU1077. Strong positive bands of desired sizes were observed for all the three mutants indicating the integration of the construct into the *PbAP2-G* locus knocking it out. The bigger sized bands appear faint as Taq polymerase in the reaction might have exhausted during its amplification thus generating weaker bands (Selectable marker cassette: orange=3'UTR of *Pbdhfr*; dark blue=*pbeef1aa*; violet=*hdhfr*; green=*yfcu*).

### 5.3.1.2 The double-crossover (recombineering) full length *AP2-G* orf knockout mutant

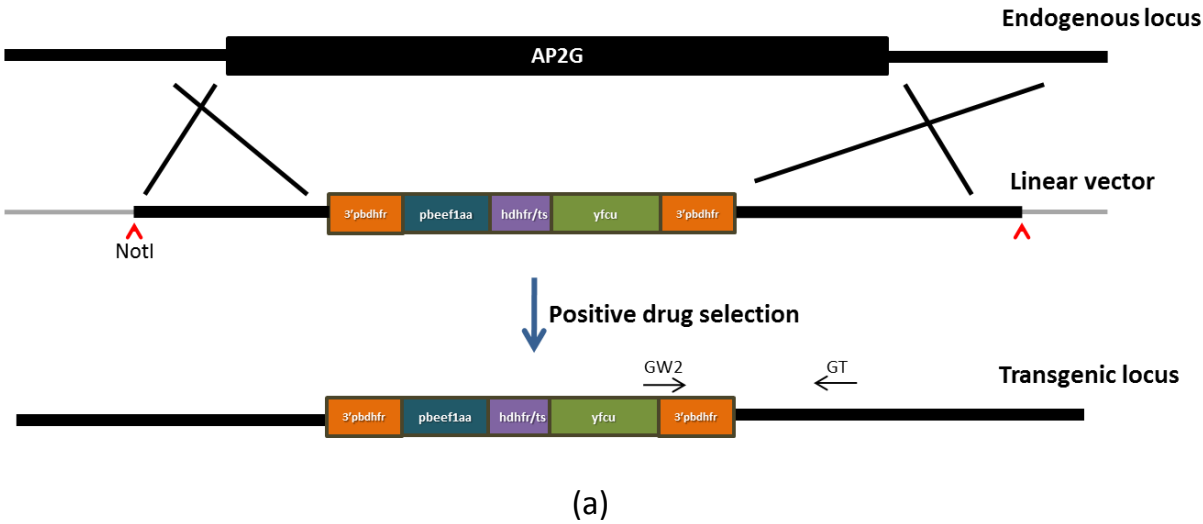
NotI-digested (New England BioLabs) and purified targeted recombineering vector DNA (pG0277) was transfected (~2 µg) to the purified schizonts of the WT line (820em1dcl2) in triplicate. The mutant lines (G355, G356 and G357) were selected by using pyrimethamine in drinking water.

Diagnostic PCRs were performed (except on G355 as the mouse unfortunately died on procedure) using oligonucleotide pairs GW2/GT (for 3' integration) as shown in the Figure 5-3. The expected length of the amplicons for the integration event was 2.4 kb (Table 5-2). None of the lines were cloned, as a cloned line of the other knockout variant (conventional DXO full orf ko) was already generated and selected for characterization. Diagnostic PCRs (Figure 5-3b) suggested successful integration of the drug selectable marker cassette at the *AP2-G* locus.

Table 5-2

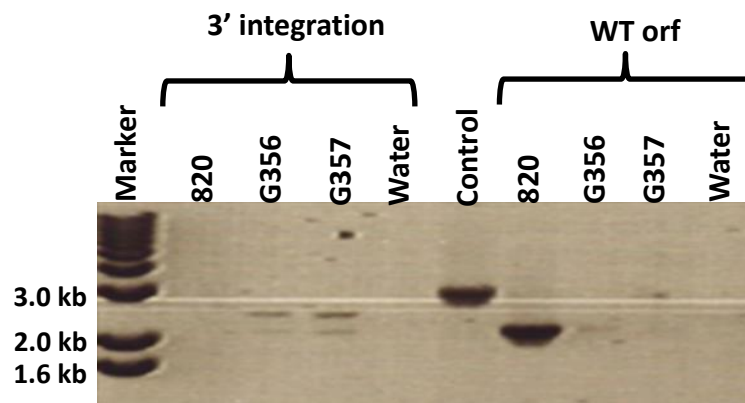
Oligo set	Details	Expected band size with DNA (kb)			Ta (°C)	Ext. Time (min)
		820	G356-7	Water		
GW2/GT	Genotyping oligos	--	2.4	--	47	5
1237/1071	Within <i>AP2-G</i> orf	2.1	--	--	54	5
1233/1242	Control (HP)	2.9	--	--	47	5

Table 5-2. Basic PCR reaction settings for amplifying various regions of the *PbAP2-G* genomic locus to identify the presence of 3' integration of the recombineering knockout construct into the genome and also to assess the presence of any wild type contamination in the knockouts.



**Figure 5-3 (a). Schematic of *AP2-G* (PBANKA\_143750) recombineering-based DXO knockout strategy (not-to-scale).** The figure shows endogenous & modified genetic locus together with the plasmid which was linearized with NotI before transfection in duplicate into the WT 820 line. The resulting transgenic mutants (G356, G357) were positively selected using the drug pyrimethamine. (Selectable marker cassette: orange=3'UTR of *Pbdhfr*; dark blue=*pbeef1aa*; violet=*hdhfr*; green=*yfcu*),





(b)

**Figure 5-3 (b). Diagnostic PCR for integration.** The PCR confirmed successful integration using the GW2/GT primers producing the 3' integration bands of 2.4 kb in each knockout. WT orf (2.1 kb) was minimally amplified from the knockout lines – this might be expected as the knockout lines were still not cloned.

### 5.3.1.3 The double-crossover (conventional) partial length *AP2-G* orf knockout mutant

HindIII and EcoRI double-digested (New England BioLabs), linearized and purified targeted vector DNA (pG0274) was transfected (~5 µg) into the purified schizonts of the WT line (820em1dcl2) in duplicate. The mutant lines (G528 and G529) were selected by using pyrimethamine in drinking water. Equal amount of purified DNA (~100 ng) from each of these mutant lines was PCR-screened for successful integration of the knockout construct into the genome by using synthesised oligonucleotide pairs GU1407/GU414 (for 5' integration) and GU392/GU1826 (for 3' integration) as shown in Figure 5-4(a). The expected lengths of the amplicons for the integration events were 1.2 kb and 1.8 kb, respectively (Table 5-3). The transfectants were also screened by amplifying for the full length of the region knocked out by using oligonucleotide pairs GU1407/GU1826 (expected band size 5.3kb in the ko mutants and 5.9 kb in the WT). Amplified DNA (25 µl) was fractionated on 1% agarose gel electrophoresis at 100V for 90 minutes to identify the bands of appropriate sizes (Figure 5-4(b)).

The integration of the construct at the 5' end could not be validated by the PCR screen as there were no bands seen. The presence of strong 3' integration bands (in both G528 and G529) together with a weak, although appropriate sized, band (5.9 kb) in G529 suggests a successful knockout of the desired *PbAP2-G* segment in G529. The line G529 was preferentially selected for

cloning over G528 as the signal for WT contamination was reduced as evident from a weaker amplification of the orf to be knocked out.

Table 5-3

Oligo set	Details	Expected band size with DNA (kb)					Ta (°C)	Ext. Time (min)
		820	G528	G529	Water	HP		
1407/414	5'Integration	NA	1.2	1.2	-	NA	51	3
392/1826	3'Integration	NA	1.8	1.8	-	NA	51	3
1407/1826	Across integration	5.9	5.3	5.3	-	NA	51	7
1241/1242	Orf	1.9	--	--	-	NA	51	3
1241/1242	Control1 (HP)	NA	NA	NA	NA	1.9	51	3
1233/1235	Control2 (HP)	NA	NA	NA	NA	5.9	51	7

Table 5-3. Basic PCR reaction settings for amplifying various regions of the *PbAP2-G* genomic locus to identify the presence of 5' and 3' integration of the partial gene knockout constructs into the genome. Ta is annealing temperature.

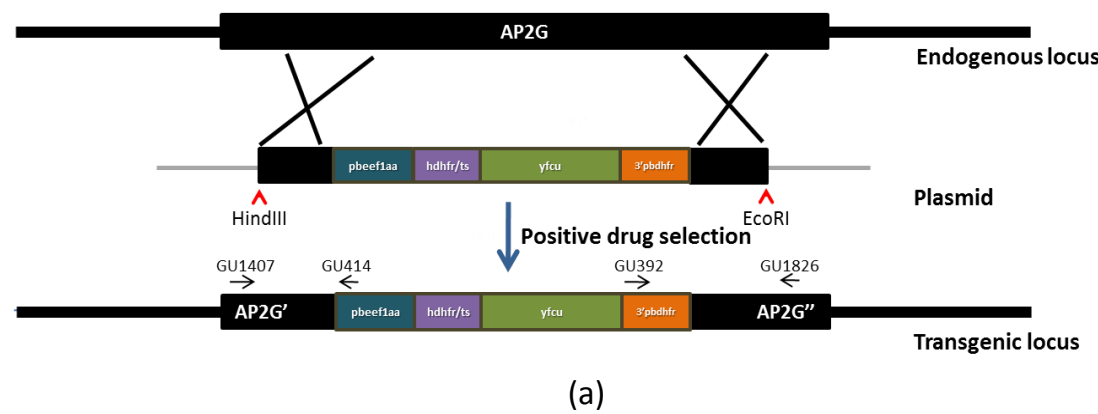
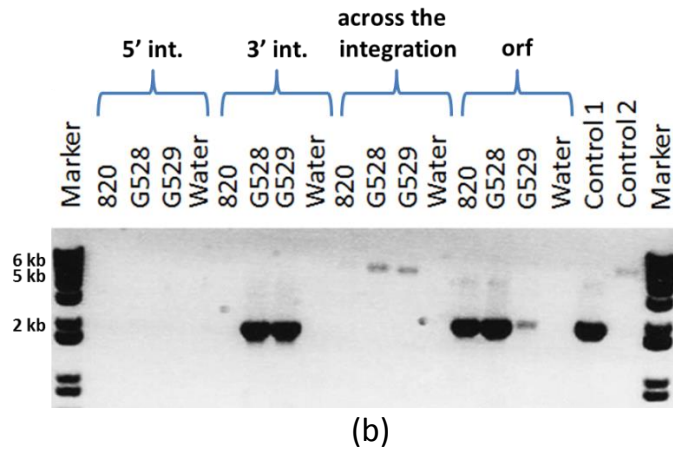


Figure 5-4 (a). Schematic of partial *AP2-G* (PBANKA\_143750) conventional DXO knockout strategy (not-to-scale) showing endogenous & modified genetic locus together with the plasmid which was linearized with HindIII and EcoRI before transfection in duplicate into the WT 820 line. The resulting transgenic mutants (G528, G529) were positively selected using the drug pyrimethamine. Also shown are the positions of the oligonucleotide primers (GU1407/GU414) and (GU392/GU1826) used for examining the successful integration of the construct into the genomic locus. (Selectable marker cassette: orange=3'UTR of *Pbdhfr*; dark blue=*pbeef1aa*; blue=*hdhfr*; green=*yfcu*).



**Figure 5-4 (b). Diagnostic PCR for integration.** The PCR confirmed successful integration using the following primers: (3' integration=1.8 kb GU392/GU1826) and (through the modified locus=5.3 kb in ko and 5.9 kb in the WT GU1407/GU1826). Strong positive bands of desired sizes were observed for the two mutants indicating the 3' integration of the construct into the *PbAP2-G* locus knocking a part of the *PbAP2-G* out.

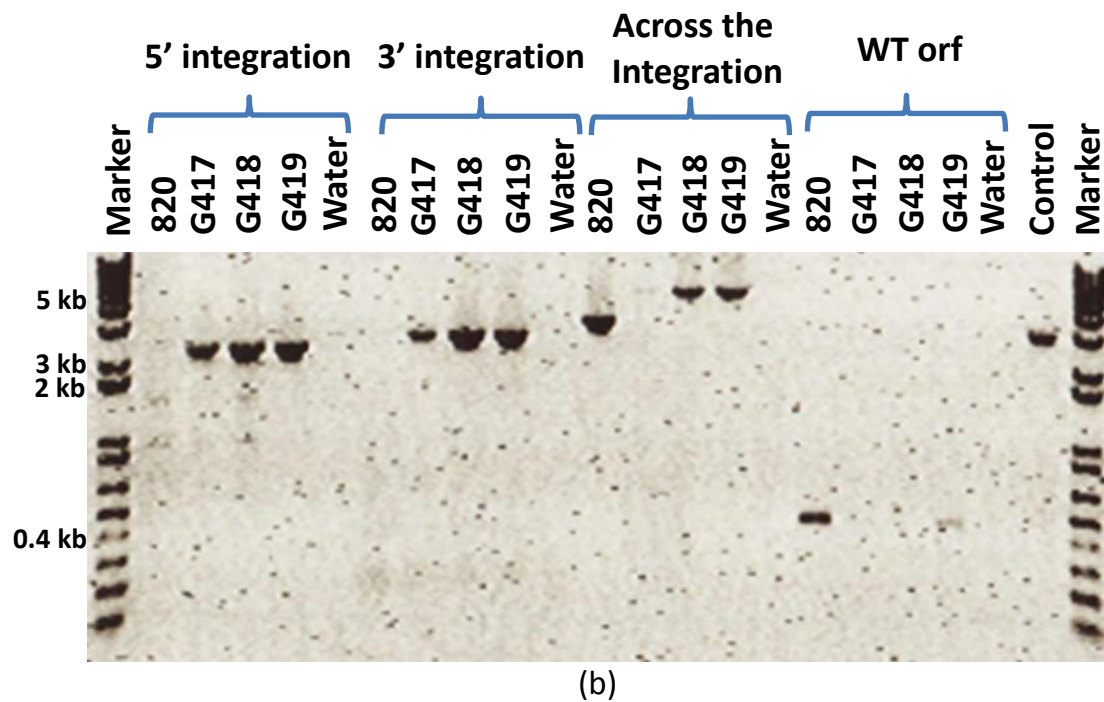
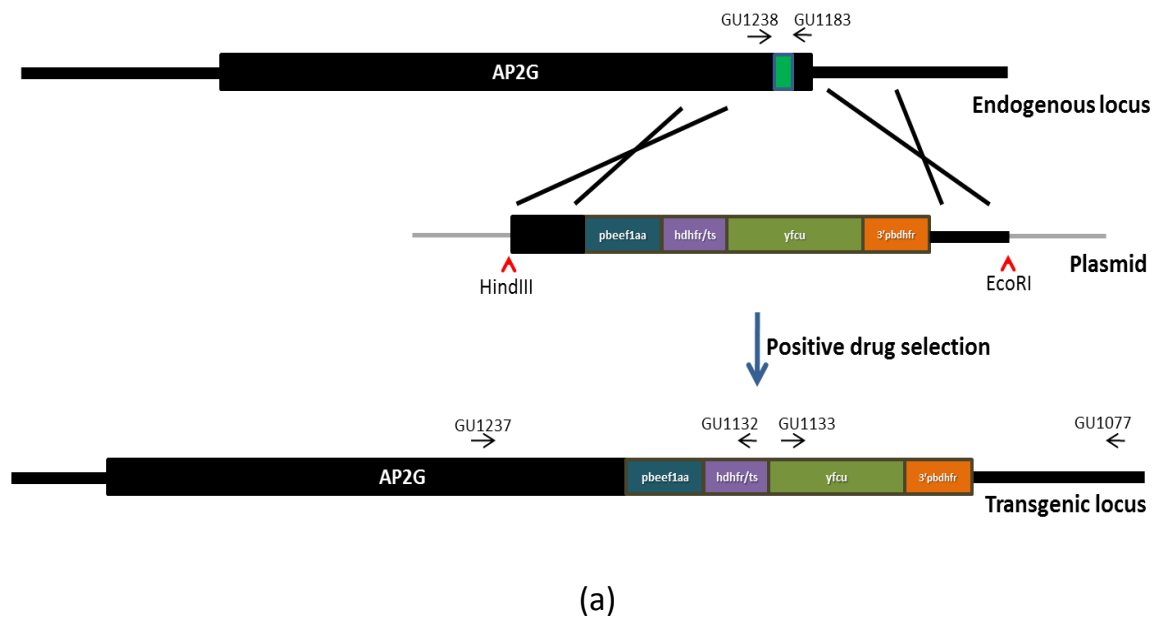
#### 5.3.1.4 The double-crossover (conventional) *AP2-G* DNA binding domain (DBD) knockout mutant

HindIII and EcoRI double-digested (New England BioLabs), linearized and purified targeted vector DNA (pG0273) was transfected (~5 µg) into the purified schizonts of the WT line (820em1dcl2) in duplicate. The mutant lines (G417, G418 and G419) were selected by using pyrimethamine in drinking water. Diagnostic PCRs (Table 5-4) using the oligonucleotide pairs GU1237/GU1132 (for 5' integration) and GU1133/GU1077 (for 3' integration) showed a successful knockout genotype (Figure 5-5). G418 was subsequently cloned successfully.

Table 5-4

Oligo set	Details	Expected band size with DNA (kb)				Ta (°C)	Ext. Time (min)
		820	G417-19	Water	HP		
1237/1132	5'Integration	--	2.2	--	NA	54	4
1133/1077	3'Integration	--	2.5	--	NA	54	4
1237/1077	Across integration	3.1	5.0	--	NA	54	7
1238/1183	Orf	0.4	--	--	NA	54	4
1182/1415	Control 1 (HP)	NA	NA	NA	4.9	54	7
1233/1242	Control 2 (HP)	NA	NA	NA	2.9	54	4

**Table 5-4. Basic PCR reaction settings for amplifying various regions of the *PbAP2-G* genomic locus to identify the presence of 5' and 3' integration of the *PbAP2-G* DBD knockout constructs into the genome.** Ta is the annealing temperature.



**Figure 5-5 (a). Schematic of *AP2-G* (PBANKA\_143750) DNA-binding domain (DBD) knockout strategy (not-to-scale).** The schematic shows endogenous & modified genetic loci together with the plasmid which was linearized with HindIII & EcoRI before transfection in triplicate into the WT 820 line. The resulting transgenic mutants (G417-9) were positively selected using the drug pyrimethamine. Also shown are the positions of the oligonucleotide primers GU1237/GU1132, and GU1133/GU1077 used for examining the successful integration of the construct into the genomic locus. **(b) Diagnostic PCR** which confirmed successful integration using the following primers: (1; 5' integration=2.2 kb) GU1237/GU1132 (2; 3' integration=2.5 kb) GU1133/GU1077 and (3; through the modified locus=5.0 kb) GU1237/GU1077. (Selectable marker cassette: orange=3'UTR of Pbdhfr; dark blue=pbeef1aa; violet=hdhfr; green=yfcu)

## 5.3.2 ***Southern analyses of cloned knockout mutants***

### 5.3.2.1 **Cloning of *AP2-G* knockout variants**

One of each of the *AP2-G* length variant knockout mutants which had the least WT contamination on PCR analyses was selected for cloning by limiting dilution method (Please see Section 2.2.11 of Materials & Methods). The clones were again analysed by PCRs for integration and WT contamination and the clone which showed no WT contamination on the PCRs for each of the knockout variants was selected for Southern analyses and subsequent experimentation.

Thus, G401, G529 and G418 were selected for cloning from each of the full-length, partial-length and DBD knockouts of the *AP2-G* open reading frame, respectively. Three clones each were obtained for G401 and G529 cloning and G401cl1 and G529cl2 were selected as the best clones to carry forward for Southern Analyses based on PCR screening. G418 cloning resulted into 6 positive clones but the cloning had to be repeated on the basis of some WT contamination being identified throughout the clones. A second round of cloning was performed on G418cl6 yielding four clones, out of which G418cl6cl3 was selected for further analyses and experimentation.

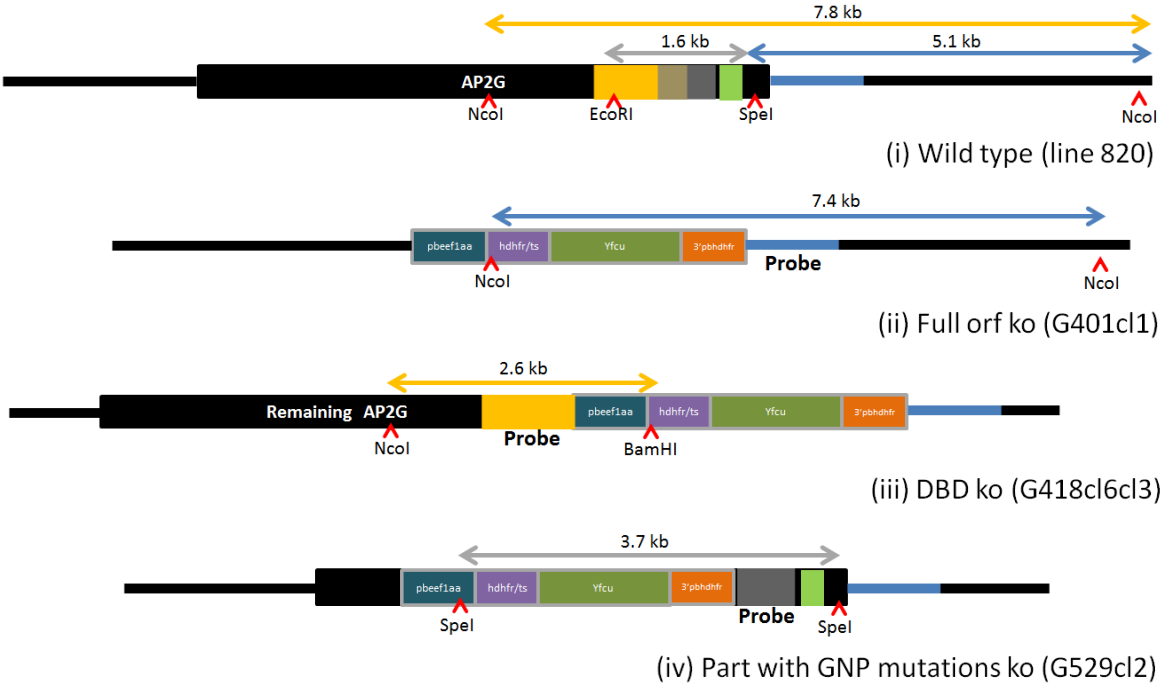
### 5.3.2.2 **Southern analyses of clones**

Technical details of Southern analysis are given in Section 2.2.4 of Materials & Methods. Briefly, for comparison with the WT line (820), purified gDNA from WT and G401cl1, WT and G418cl6cl3 & WT and G529cl2 was double-digested with the High-Fidelity (HF) versions of NcoI & SpeI, NcoI & BamHI and EcoRI & SpeI, respectively (Table 5-5 and Figure 5-6a). The membrane was then divided into three parts and then hybridized separately with P<sup>32</sup> labelled single-stranded DNA probes from specific regions from one of the homology arms used for generating the gene targeting vector. The probes were PCR-amplified and purified using the following oligonucleotides: GU1058/GU1059 for G401cl1, GU1416/GU1417 for G418cl6cl3 and GU1414/GU1415 for G529cl2. Each of the three membranes was washed and exposed for 35 hours (Figure 5-6b).

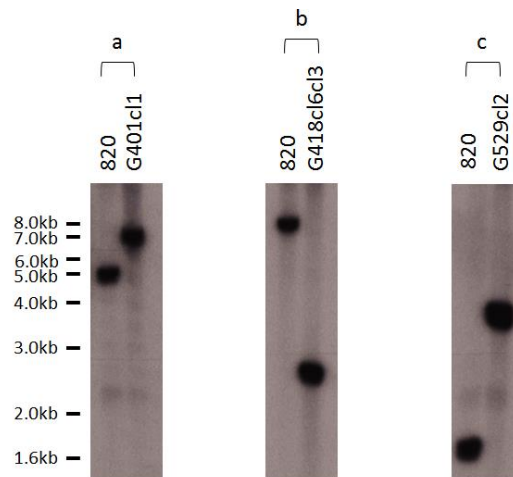
Table 5-5

	WT and G401cl1			WT and G418cl6cl3			WT and G529cl2		
	WT	G401cl1	Plasmid	WT	G418cl6cl3	Plasmid	WT	G529cl2	Plasmid
RE	NcoI & SpeI			NcoI & BamHI			EcoRI & SpeI		
Probe	Downstream homology arm of full orf ko construct			Upstream homology arm of the DBD ko construct			Downstream homology arm of part orf ko construct		
Band size (kb)	5.1	7.4	7.2	7.8	2.6	6.0	1.6	3.7	3.4

**Table 5-5. Restriction digest pattern and probes used for Southern analyses of the three *PbAP2-G* cloned knockout variants.** Genomic DNA from WT & G401cl1, WT & G418cl6cl3 and WT & G529cl2 were double-digested with NcoI/SpeI, NcoI/BamHI and EcoRI/SpeI, respectively. The differential restriction digest pattern in each of the WT/knockout duo as evident by the difference in the size of the observed bands in each digest was clearly visible (Figure 5-6b).



**Figure 5-6 (a): Southern blot: Schematic (not to scale) showing the restriction digestion pattern of gDNA from *AP2-G* full orf knockout (ii), *AP2-G* DBD knockout (iii) and *AP2-G* part orf bearing the GNP mutations 7, 8,9 knockout (iv) double-digested with NcoI/SpeI, NcoI/BamHI and EcoRI/SpeI, respectively.** The restriction pattern of WT 820 line being digested with all the 4 enzymes is shown in (i). The red arrowheads denote the site of action of the respective RE's. The digested DNA was fractionated in 0.7% agarose and transferred to a nylon membrane. The membrane was UV-cross-linked and cut into three and subsequently probed separately with  $P^{32}$  labelled single-stranded PCR-amplified probes derived from the following homology arms: downstream homology arm for (ii) shown as blue bar immediately downstream to the *AP2-G* orf, upstream homology arm for (iii) shown as yellow bar within the *AP2-G* orf and downstream homology arm for (iv) shown as grey striped bar within the *AP2-G* orf. (Selectable marker cassette: orange=3'UTR of Pbdhfr; dark blue=pbeef1aa; violet=hdhfr; green=yfcu).



**Figure 5-6 (b): Southern blots showing the restriction digestion pattern of *NcoI/SpeI* digested gDNA (a) from 820 WT (5.1 kb) and G401cl1 (7.4 kb); *NcoI/BamHI* digested gDNA (b) from 820 WT (7.8 kb) and G418cl6cl3 (2.6 kb); and *EcoRI/SpeI* digested gDNA (c) from 820 WT (1.6 kb) and G529cl2 (3.7 kb).**

## 5.4 The knockout phenotypes

The effect of different length variant knockouts of the *AP2-G* gene was assessed on the basis of the resulting phenotypic changes on the sexual stage of development of *Plasmodium*. The modalities to test this included Giemsa-stained smears from peripheral blood, flow cytometric (FACS) analyses of the mutant lines, exflagellation assays, ookinete cultures, transmitting the mutant lines through mosquitoes, and northern and RT-PCR based analyses of the knockout mutants.

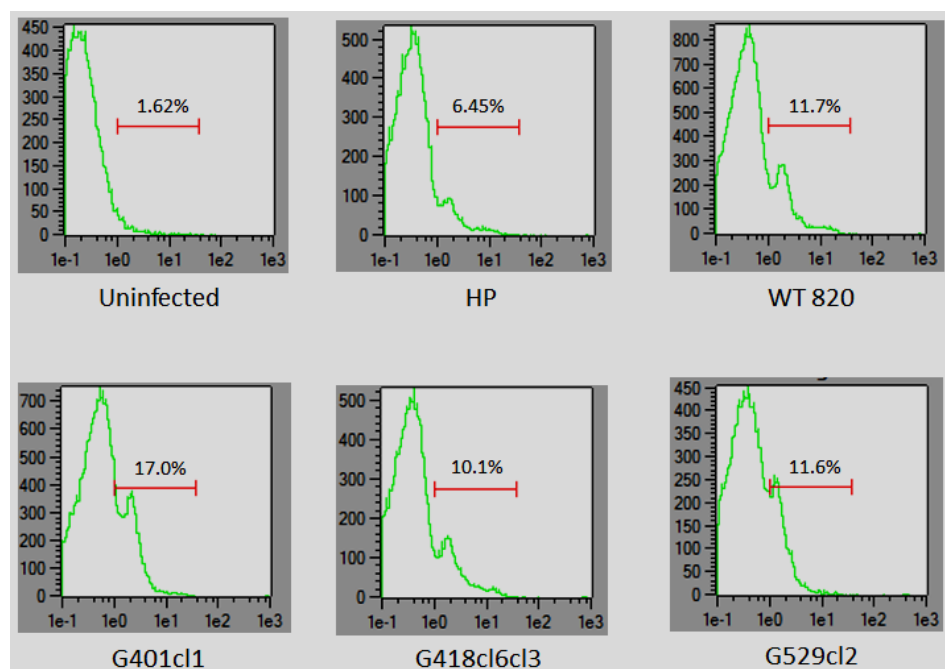
### 5.4.1 *Giemsa-stained peripheral smears*

Peripheral blood smears from the cloned knockout lines and WT line 820 (as control) at total parasitaemia between 5% and 10% were stained with Giemsa stain and examined for gametocytes under a microscope (please see Section 2.1.3 of Materials & Methods for details). Fifteen hundred parasites for each of the mutant and WT lines were examined and the number of gametocytes observed recorded as a percentage of total parasites. No gametocytes could be identified in all three knockout variants as compared to 11 % gametocytes in the WT line 820. These results strongly favoured absence of gametocytemia in all the *AP2-G* length variant knockout lines.



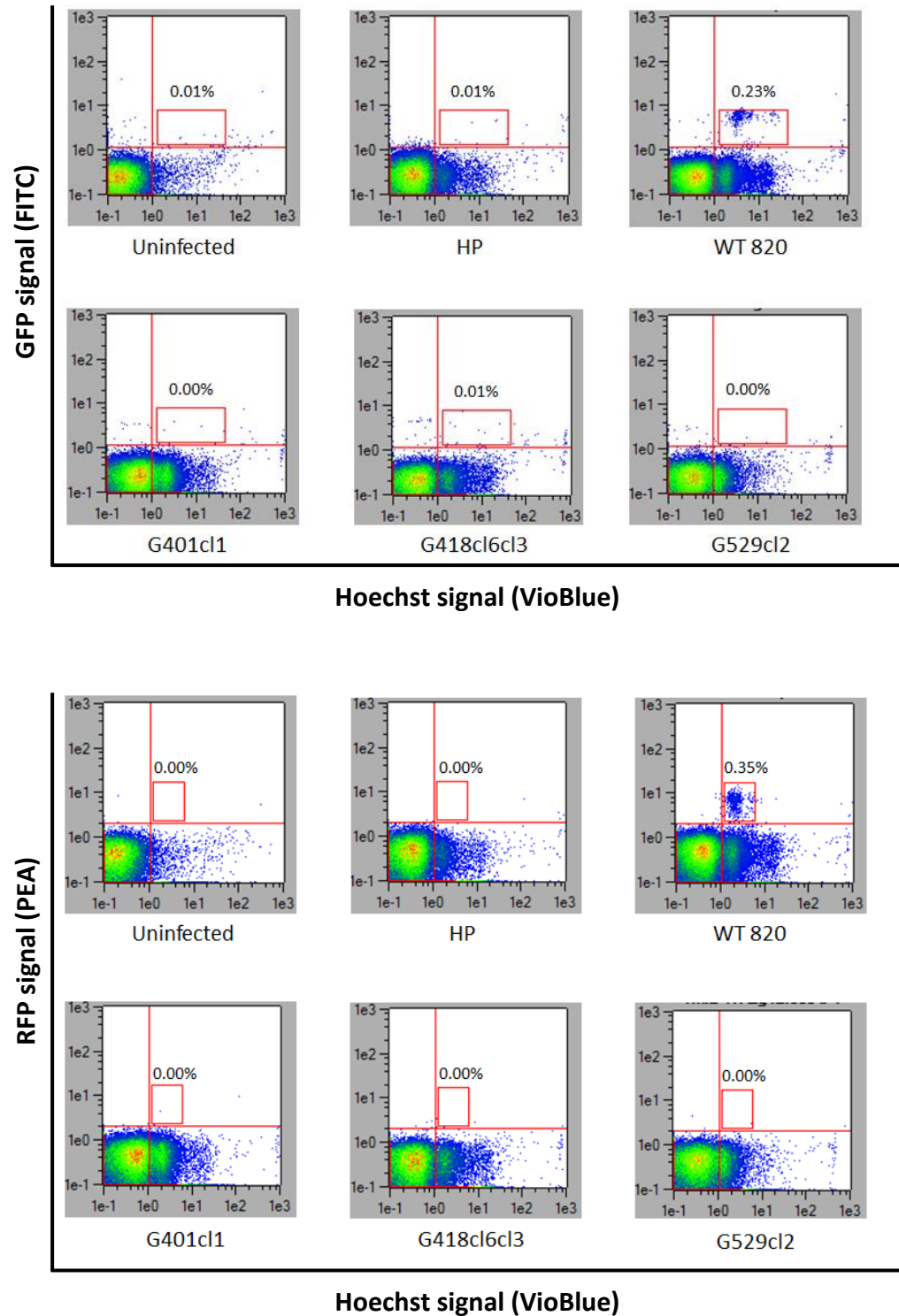
### 5.4.2 *Flow cytometry-based evidence*

Hoechst-stained peripheral blood from the cloned knockout lines, along with WT line 820 (as positive control), non-fluorescent high gametocyte producer line HP (as negative control) and uninfected mouse (also as negative control) was examined using flow cytometry (please see Section 2.5 of Materials & Methods for details) as in Figure 5-7. Samples were examined when the total parasitemia in the infected lines reached between 5% and 10%. Convincingly, none of the mutant lines showed any gametocytemia, further indicating that these lines no longer produce any gametocytes.



**Figure 5-7 (a). FACS plots showing the Hoechst signals detected in the VioBlue channel in all six lines examined.** VioBlue signals (the green peaks), as they come from the Hoechst-stained DNA, reflect the proportion of infected red blood cells detected in the gated population. The 200,000 red blood cells/events that were examined on the Forward and Side Scatter plots from FACS (not shown here) were appropriately gated (selected) to select predominantly for the red blood cells to be included in further analyses (shown here). Thus, majority of the other potential sources of DNA have been excluded from the analyses. The reading (reflected from the green peaks) in these plots indicate the proportion of parasite-infected cells together with some background noise as detected in the sample from the uninfected mouse (negative control). The population of cells showing the VioBlue signals in these plots were again gated (red horizontal bars) to select for the Hoechst-positive population that best reflects the infected red cells predominantly (shown as percentage over the red bars). The percentage of infected cells (1.62%) detected from the uninfected negative control line reflects the background false positive noise and hence should be discounted from the readings obtained from other infected lines. The undiscounted parasitemias in the HP, WT 820, G401cl1, G418cl6cl3 and G529cl2 lines was 6.45%, 11.7%, 17%, 10.1% and 11.6%, respectively. The red gated population from these plots was included in further analyses of red and green fluorescence (see Figure 5-7b and 5-7c).





**Figure 5-7 (b and c).** FACS plots showing the GFP and RFP signals (Y-axis) detected in FITC and PEA channels, respectively in Figures 6-7 (b) and (c), in the Hoechst-positive population (X-axis) gated in Figure 5-7a among all the six lines examined. Each blue dot lying in the red rectangular gate represents a Hoechst-positive red blood cell which is positive for GFP/RFP signals as well, and thus is a male/female gametocyte, respectively, within the RBC. The proportion of such cells (shown as percentage above the red rectangular gate) represents the percentage of male/female gametocytes among the infected red blood cells. The male and female gametocytemia shown for the uninfected line (0.01% and 0.0%, respectively) and for the high gametocyte producer non-GFP/RFP fluorescent line HP (0.01% and 0.0%, respectively) thus reflects background fluorescence (false-positive error) and thus was discounted from the gametocytemia shown for the WT line 820 and the three cloned knockout lines. The undiscounted male gametocytemia in the WT 820, G401cl1, G418cl6cl3 and G529cl2 lines was 0.23% and 0% each, respectively. The undiscounted female gametocytemia in the WT 820, G401cl1, G418cl6cl3 and G529cl2 lines was 0.35% and 0% each, respectively.

### 5.4.3 *Exflagellation assays*

In the absence of any visible gametocytes on Giemsa-stained smears and flow cytometry, the presence of any undetected functional male gametocytes in the knockout lines was investigated using the capacity of male gametocytes to exflagellate *in vitro* in ookinete culture media. The male gametocytes from peripheral blood, if any, in the WT line 820 and in each of the knockout variants were allowed to undergo the process of exflagellation reaction in the ookinete culture media. The number of exflagellation centres (ECs) formed in the WT line was then compared to in each of the knockout variants and expressed as ECs per  $10^4$  red blood cells examined. No ECs were observed in any of the *AP2-G* knockout variants in comparison to 29 ECs in the parental WT line 820, further strengthening the evidence that none of these mutants is gametocytogenic (Table 5-6).

### 5.4.4 *Ookinete cultures*

Overnight ookinete cultures were set up with 1-2 drops of peripheral blood from the WT line 820 and each of the knockout variants being incubated with ~5 ml of ookinete culture media at 21 °C overnight to examine the presence of ookinetes under a microscope (expressed as the number of ookinetes per 20 fields), thus documenting the presence of functional male and female gametocytes, if any. No ookinetes were identified in all the *AP2-G* knockout variants as compared to 21 ookinetes in the WT parental line 820 (Table 5-6).

Table 5-6

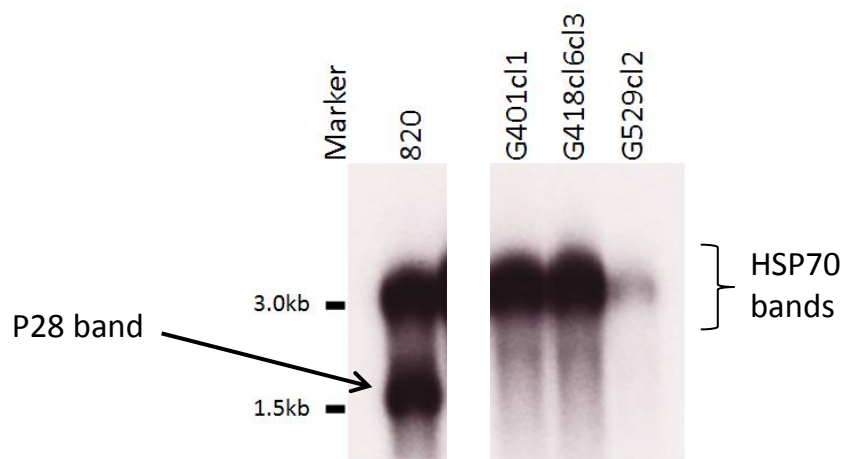
	820 (WT)	G401cl1 (full orf ko)	G529cl2 (partial orf ko)	G418cl6cl3 (DBD ko)
<b>Flowcytometry (FACS)</b> (percentage gametocytes out of the total infected RBCs)	<b>0.58</b>	<b>Zero</b>	<b>Zero</b>	<b>Zero</b>
<b>Giemsa-stained smears</b> (% of gametocytes per 1500 parasites)	<b>11%</b>	<b>Zero</b>	<b>Zero</b>	<b>Zero</b>
<b>Exflagellation assay</b> (number of Exflagellation Centres per $10^4$ RBCs; n=3)	<b>29.6</b>	<b>Zero</b>	<b>Zero</b>	<b>Zero</b>
<b>Ookinete counts</b> (number of ookinetes per 20 fields)	<b>21</b>	<b>Zero</b>	<b>Zero</b>	<b>Zero</b>
<b>Mosquito passage</b> (Transmission bite-back)	<b>Parasites seen</b>	<b>No parasites</b>	<b>No Parasites</b>	<b>No Parasites</b>

**Table 5-6. A composite table showing the examined phenotype of the three knockout variants of *PbAP2-G* as compared to the WT parental line 820.** No evidence of gametocytes being produced by any of the three *PbAP2-G* knockout variants (G401cl1 – full orf ko; G529cl2 – partial orf ko; and G418cl6cl3 – DBD ko) as compared to the WT parental line 820 could be established by FACS (No observable gametocytes in the three ko's compared to 11% in the WT), Giemsa-stained smears (No observable gametocytes per 1500 parasites in the three ko's compared to 0.58% in the WT), Exflagellation assay (No observable exflagellation centres (EC's) per  $10^4$  RBCs in the three ko's compared to 29.6 EC's in the WT), Ookinete cultures (No observable ookinete per 20 fields in the three ko's compared to 21 in the WT), and finally mosquito passage (No parasites in the three ko's transmission bite-back mice compared to 0.58% in the WT). All these evidences taken together convincingly demonstrate the absence of gametocytogenesis in all the knockout mutants.

### 5.4.5 Northern analyses of the mutants

After ruling out the presence of any sexual stage life forms of *Plasmodium* in all the *AP2-G* knockouts, the expression of the sexual stage specific transcript (*P28*) was examined using the northern blot technique. Briefly, approximately 5

µg of RNA sample for each of the WT parental line 820 and the three knockout variant lines (except G529cl2; which was ~2 µg) was used for the northern analyses. The RNA on the membrane was probed for the sexual stage transcript with P<sup>32</sup> labelled *P28* DNA probe (PBANKA\_051490; 0.62 kb cds) and normalised using an HSP70 probe (PBANKA\_071190; 2.08 kb cds). The absence of *P28* signals from all the *AP2-G* knockouts is consistent with the absence of sexual stages in these lines (Figure 5-8). Northern analyses of the three knockouts were performed together with the three de novo GNP mutants. The gel picture in Figure 5-8 has been cropped to remove the three lanes bearing the natural GNP mutants intervening between the WT 820 lane and the lanes bearing the three knockouts.

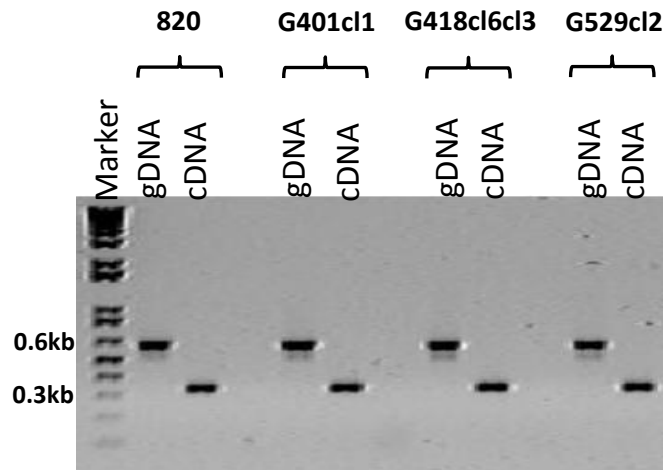


**Figure 5-8. Northern blot of total RNA from lines 820 (WT line) and *AP2-G* knockout length variant mutant lines G401cl1 (full orf knockout), G418cl6cl3 (DNA binding domain knockout) and G529cl2 (partial orf knockout eliminating the region of the orf with the three natural GNP mutations).** The blot was probed for sexual stage specific transcripts with <sup>32</sup>P labelled *P28* dsDNA (~1.5 kb band in WT line 820 and no bands in the ko lines) and normalized with HSP70 probe (PBANKA\_071190; ~3 kb band in all lines). The absence of *P28* signal from all the ko lines is highly suggestive of absence of sexual stages in the GNP lines. (Note: the original blot had three extra lanes for showing the absence of *P28* signals in the three naturally acquired GNP lines after the lane for the WT 820 line; the same have been cropped out of the image (and shown elsewhere in Chapter 3; Figure 3-17) and the rest of the image is shown here as two images but they come from the same gel).

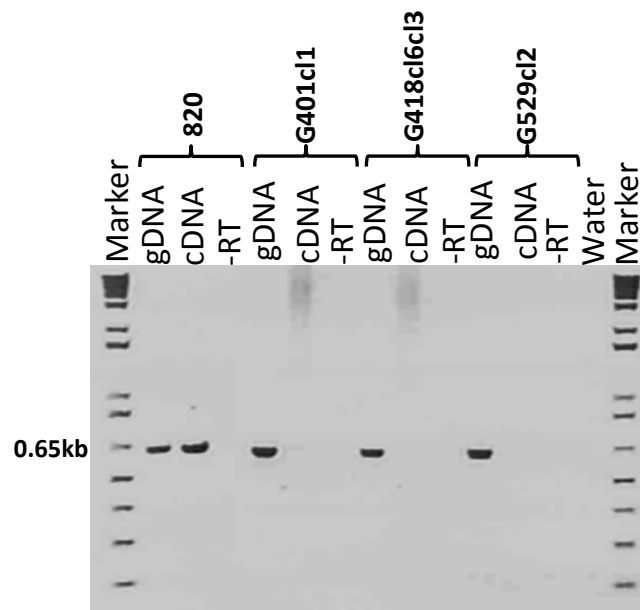
#### 5.4.6 RT-PCR analyses

The presence of any sexual stage-specific transcripts for *P28* was again ruled out using reverse-transcriptase PCR (RT-PCR). Briefly ~4 µg of purified DNase treated total RNA from the WT parental line 820 and the three *AP2-G* knockout variants was subjected to RT-PCR, as mentioned under Section 2.2.6 of

**Materials & Methods.** The efficiency of the RT reaction (and hence contamination of the cDNA with the gDNA, if any) was tested by amplifying the constitutive *ALBA3* (PBANKA\_120440) from all the lines tested using primer sets GU515/GU516. Since *ALBA3* gDNA has an intron, its PCR amplification after the RT reaction allowed differentiation between amplicons from gDNA and cDNA on the basis of difference in expected sizes: 0.64 kb for gDNA and 0.32 kb for cDNA.



(a)



(b)

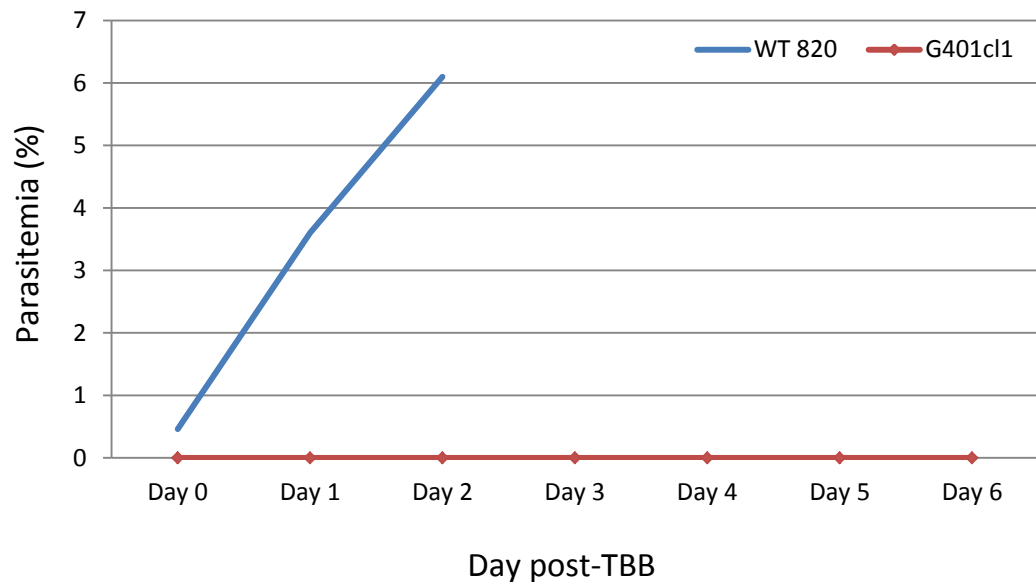
**Figure 5-9. Agarose gel electrophoresis pictures showing amplified bands from RT-PCR reactions of *P. berghei* *ALBA3* (constitutive) and *P. berghei* *P28* (sexual stage-specific).** Figure 5-9 (a) shows the *ALBA3* gDNA and *ALBA3* cDNA (RT-PCR products) from the WT parental line 820 and the three *PbAP2-G* knockouts to assess the efficiency of the reverse transcriptase reaction in terms of gDNA contamination of cDNA. Since the *alba3* gene (0.6 kb) contains an intron (0.3 kb) provides a basis of detecting the gDNA (0.6 kb band) contamination of the cDNA. Absence of any 0.6 kb bands in all the samples rules out the possibility of the presence of any gDNA in the RT-PCR products. Sexual stage-specific *P28* was reverse-transcribed and PCR-amplified from the WT line 820 but not the three *PbAP2-G* knockout variants Figure 5-9 (b).

Presence of *P28* gDNA bands in all the lines and *P28* cDNA only in the parental WT 820 line together with similar findings from the northern analyses strongly establishes the absence of any expression of sexual stage-specific prototype *P28* from all knockout lines, thus supporting an absence of sexual stages in the knockouts.

The RT reaction yielded pure cDNA (uncontaminated with any gDNA) as evidenced by the presence of the smaller 0.32 kb band and absence of the gDNA band in the *ALBA3* RT-PCR amplicons (cDNA). Absence of the sexual stage specific *P28* cDNA band in the RT-PCR analyses from all three *AP2-G* knockout lines (in presence of the *P28* band in gDNA and cDNA from WT 820 line) further corroborated the findings of the northern analyses, strongly indicating an absence of sexual stages in the three *AP2-G* knockout variants. No bands in the RT reactions ruled out the presence of any gDNA contamination in the samples (Figure 5-9b).

#### **5.4.7      *Transmission through mosquitoes***

As a final confirmation of the absence of any residual functional gametocytes in the three *AP2-G* knockout variants, the full length *AP2-G* knockout variant (G401cl1) was transmitted through the mosquitoes along with the WT parental line 820 in separate experiments. Briefly, a mouse infected independently with the WT line 820 was fed to approximately 200 adult mosquitoes (predominantly females) and a mouse infected with G401cl1 was fed to approximately 600 adult mosquitoes (predominantly females) in independent cages at a parasitemia between 2% and 4%. The mosquitoes were maintained at 21 °C for 21 days in the insectary after which they were fed on naïve mice (mosquito bite-back or transmission bite-back or TBB). Blood samples from the mice were smeared from day 0 of the bite-back and were examined for parasites/gametocytes by Giemsa stain and flow cytometry (FACS). The WT line 820 bite-back mouse had detectable parasitemia in the peripheral blood on day 0 post-bite back and the parasites followed normal growth kinetics. No parasitemia could be detected in G401cl1 TBB line up to day 6 post-TBB (Figure 5-10) and even on day 10 and day 14 of the bite-back (data not shown), confirming that the cloned knockout line lacked viable gametocytes when initially fed by the mosquitoes.



**Figure 5-10. Post-transmission parasitemia kinetics as percentage of infected red blood cells in lines WT 820 and G401cl1 (full orf ko line) on Y-axis and days post-biteback on X-axis.** WT 820 line showed detectable parasitemia straightaway from the day the infected mosquitoes were allowed to feed on the naïve mice. The knockout line G401cl1 did not show any parasites up to day 6 post-TBB and even by day 14 post-TBB (not shown in the kinetics above). These findings show conclusive evidence that the knockout line, G401cl1, failed to transmit through the mosquitoes as they lacked functional gametocytes.

## 5.5 Summary and discussion

Gametocytogenesis in *Plasmodium* has been poorly understood in terms of the genetic regulation of the process in general and the molecular basis of commitment to gametocytogenesis in particular (Eksi et al., 2012; Ikadai et al., 2013). Although microarray experiments in *P. falciparum* have identified many genes being associated with early and mature gametocytes as gametocyte-specific genes (Young et al., 2005), targeted gene knockout of some of them including *Pfs16* (Kongkasuriyachai et al., 2004), *Pfg27* (Olivieri et al., 2009), *Pfmdv1/peg3* (Furuya et al., 2005), *Pf11.1*, *Pfs230* (Eksi et al., 2006) and *Pfg377* (de Koning-Ward et al., 2008) did not produce an absolute gametocyte non-producer mutant.

The role of ApiAP2 TF PBANKA\_143750 (termed AP2-G) in commitment to gametocytogenesis (established earlier in Chapter 3 and 4) was tested by a reverse genetics approach, using conventional and recombineering-based homologous recombination, by knocking out the *PbAP2-G*. Because of the relatively large size of the gene, at least two different knock out strategies and

knock out involving three different regions of the *AP2-G* gene (including the complete open reading frame) were attempted. A variety of experiments targeted towards understanding the phenotype of the three length-variants of *P. berghei* *AP2-G* (full length, partial length and DNA-binding domain) knockout convincingly established a central role for the transcription factor in gametocytogenesis.

Knockout of the implicated gene and correlation of the resulting phenotype with the expected role played by the gene are amongst the first steps towards establishing a causative association of the implicated gene with the phenotype. *AP2-G* was implicated to play a central role in commitment of the parasite towards sexual development on the basis of its uniqueness in being mutated in all three *de novo* GNP mutants generated in this study. Three independent *Pb AP2-G* targeted knockout constructs varying in the length of the gene region to be knocked out were made including: a full length *AP2-G* knockout, a partial length *AP2-G* knockout deleting the region which had the three GNP mutations and a DBD knockout. Each construct was transfected in triplicate to generate three mutant lines from each knockout variant.

The *PbAP2-G* knockout length variants also established the criticality of different regions of the protein in terms of their functional requirement for gametocytogenesis. Whereas the full-length and DBD knockout associated GNP phenotype revealed the expected functional essentiality of the domain, the partial length knockout variants eliminating the portion of the orf bearing the three GNP mutations explained the contributory role of frameshift and premature stop codons upstream to the DBD in gene expression. However, it would still be interesting to examine the role of the region downstream to the DBD as it was involved in the GNP attributed to the mutation in line 233.

Collective evidence from the genotypic and phenotypic analyses of the knockout variants unambiguously revealed that none of these mutants produced viable gametocytes thus confirming the central role of *PbAP2-G* in commitment to *Plasmodium* gametocytogenesis. Independent experiments with the *P. falciparum* ortholog of PBANKA\_143750, PF11\_1085w knockout also identified a GNP phenotype (Kafsack et al., 2014). This conservation of function of the *AP2-G*



gene in both the species is indicative of a conserved, critical role of the gene in the process of commitment to gametocytogenesis. Recent work also demonstrated the role of another ApiAP2 gene, PBANKA\_103430 (termed *AP2-G2*), in gametocytogenesis. In an independent work, disruption of *AP2-G2* resulted in near-complete loss of production of mature gametocytes (a few female gametocytes occasionally observed) and failure to transmit through mosquitoes (Sinha et al., 2014; data not shown). Direct growth competition assays between *AP2-G* knockout, WT *P. berghei* and *AP2-G2* knockout revealed that *AP2-G* knockout mutants overgrew the WT and *AP2-G2* knockout mutants indicating that *AP2-G* mutants completely lacked gametocytes thus offering a growth advantage to the asexually dividing parasites (Hughes and Billker, personal communication). This establishes the central role of *AP2-G* at the point of commitment to gametocytogenesis whereas *AP2-G2* is probably involved downstream of *AP2-G* once the parasites are committed to form gametocytes.

**6 Gene complementation in gametocyte non-producer mutants to restore their commitment to gametocytogenesis**

## 6.1 Introduction

A forward genetics whole genome sequencing-based approach to identify the genetic locus involved the generation of GNP mutants (Chapters 3 and 4) together with the targeted gene perturbation studies (Chapter 5) is generally sufficient to establish causality between genotype and phenotype. However for complete rigour it was decided to attempt complementation of the implicated gene to restore the WT phenotype. Plus, it was also thought that complementation would open the possibility to attempt to complement other GNP lines that have *AP2-G2* defects in an attempt to ascertain if they were now capable of transmission.

Genetic complementation opens opportunities to causally associate an implicated gene with its hypothesised function. Complementation experiments may serve as powerful tools in understanding the molecular basis of a variety of basic and applied phenomena in apicomplexan biology where most of the genes have hypothetical and/or unknown functions (Striepen et al., 2001; Sultan et al., 2001; Florent et al., 2010). Also, even in the presence of the usual controls, it may be difficult to refute that the resulting phenotype from gene targeting experiments is not due to accumulation of independent genomic alterations during the recombination events or due to the presence of the altered recombinant / genomic locus *per se* (Janse et al., 2011).

However, difficulty in cloning large coding regions of AT-rich genomes in order to be able to complement the full length of the knocked out gene still remains a major deterrent to successful complementation experiments in *Plasmodium spp.*, more so when the region to be complemented extends up to not less than 7 kb, as in the case here with *PbAP2-G*. The role of a second independent mutant is said to be indispensable in difficult to complement genes (Goldberg et al., 2011).

Amidst all the difficulties in attempting successful complementation in malaria parasites, especially *P. berghei*, the novel Gene-In Marker-Out, (the GIMO) approach claims to offer multiple advantages over conventional complementation approaches in terms of being faster, less laborious and

requires fewer animals (Lin and Annoura et al., 2011). Additionally, the GIMO method could be combined with the recently developed “recombineering” based high-throughput gene targeting approach (Pfander et al., 2011).

## **6.2 Complementation strategies to restore gametocytogenesis in the *PbAP2-G* mutants**

As two different types of *PbAP2-G* mutants were created using forward and reverse genetic approaches independently, both generating the gametocyte non-producer (GNP) phenotype, restoring the full functional copy of *PbAP2-G* was attempted through at least two independent approaches. Whereas forward genetic experiments identified three independent natural GNP *PbAP2-G* mutants (having SNPs and indels; Chapter 3), the reverse genetics approach resulted in creating three knockout length variants of *PbAP2-G* in separate experiments (Chapter 6). As the two approaches resulted in a differently modified *PbAP2-G* locus, complementation strategies differed for each group of GNP mutants.

### **6.2.1 Complementation strategies for natural *PbAP2-G* mutants**

Chapter 3 of this thesis identified three natural GNP mutants (m7, m8 and m9) by forward genetic screening all these GNP mutants had a unique and therefore independent mutation in *PbAP2-G* - SNP in m7 and indels in m8 and m9. As these GNP mutants were negatively selected to remove the drug selectable markers initially in their genome, complementation was attempted with constructs having a full length *PbAP2-G* (~7 kb) flanked either by an upstream (PbG01\_COMP-UP) or a downstream (PbG01\_COMP-DOWN) copy of the positive-negative drug selectable marker.

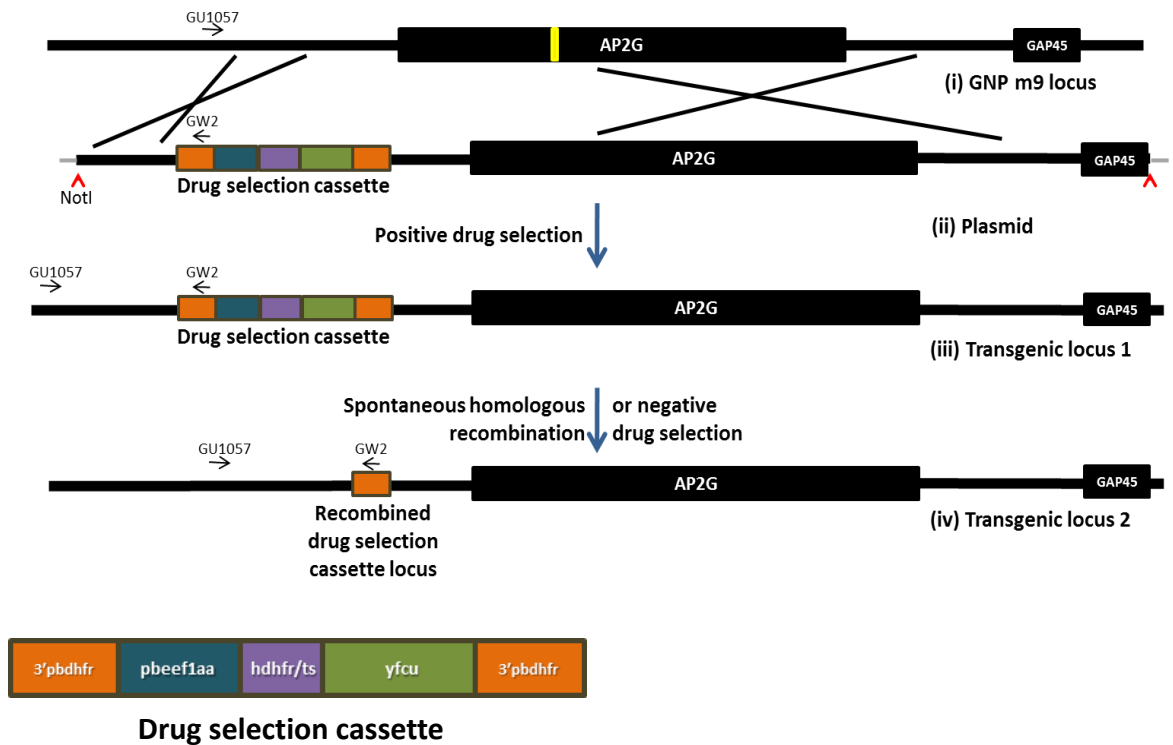
### 6.2.1.1 Complementation of natural *PbAP2-G* GNP mutants with PbG01\_COMP-UP

In order to complement all three natural *PbAP2-G* GNP mutants with a full length copy of the WT *PbAP2-G*, a recombineering based construct (PbG01\_COMP-UP; Figure 6-1a) was kindly designed and gifted by Claudia Pfander and Oliver Billker (Wellcome Trust Sanger Institute, Cambridge, UK). The construct had a full length copy of the WT *PbAP2-G* with a positive-negative drug selectable marker (positioned 1.2 kb upstream of the *AP2-G* translation start site) cloned into a linear pJAZZ vector. The purified vector was double-digested with NotI (New England BioLabs) and approximately 5 µg of the linearized vector was transfected into the cloned GNP m9 line (GNPm9cl1 as a prototype) in triplicate. The resulting transfectants (G366, G367 and G368) were selected by using pyrimethamine in drinking water. Transfectants were PCR-screened for successful 5' integration by using synthesised oligonucleotide pairs GW2/GU1057 as shown in the Figure 6-1(a). The expected length of the amplicons for each of the integration events was 1.5 kb (Table 6-1). A summary of all the oligonucleotide sets used as primers in the PCRs, their details, expected length of bands, annealing temperature (Ta) settings used and extension time used for various PCR reactions are shown in Table 6-1.

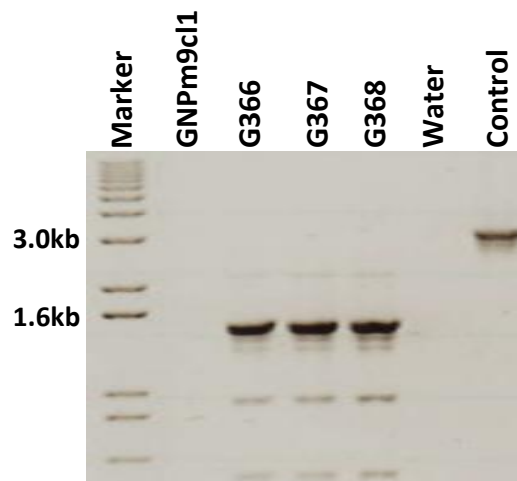
Table 6-1

Oligo set	Details	Expected band size with DNA (kb)				Ta (°C)	Ext Time (min)
		GNPm9cl1	G366-8	Water	Control		
GW2/1057	5'Integration	--	1.5	--	NA	48	5
1233/1242	Control	NA	NA	NA	2.9	48	5

**Table 6-1. Basic PCR reaction settings for amplifying various regions of the *PbAP2-G* genomic locus to identify the presence of 5' integration of the recombineering based complementation construct (PbG01 COMP-UP) into the GNP9 genome.**



(a)



(b)

**Figure 6-1 (a). Schematic of complementation strategy used to repair the m9 mutation in the *PbAP2-G* (not-to-scale).** The figure shows mutated endogenous GNP m9 & modified genetic locus together with the recombineering based plasmid construct which was linearized with *NotI* before transfection in triplicate into the cloned GNP m9 line. The endogenous GNP m9 locus (i) shows the location of the SNP in GNP m9 as a yellow bar on the *PbAP2-G* orf. The drug selectable marker (multi-coloured bar; orange=3'UTR of *Pbdhfr*; dark blue=*pbeef1aa*; violet=*hdhfr*; green=*yfcu*) is positioned at 1.2 kb upstream of the *PbAP2-G* translation start site (ii). The resulting transgenic mutants (G366-68) were positively selected using the drug pyrimethamine containing the transgenic locus (iii) with repaired m9 mutation. Also shown are the positions of the oligonucleotide primers (GU1057/GW2) used for examining the successful 5' integration of the construct into the genomic locus. The recombinogenic potential of the drug selection cassette (either spontaneous or under negative selection) and the resultant transgenic locus (iv) is also shown in the schematic. **Figure 6-1(b). Diagnostic PCR showing integration events.** The PCR confirmed successful 5' integration using the following primers GU1057/GW2 (5' integration band = 1.5 kb). Strong positive bands of desired size were observed for all the three mutants indicating the integration of the construct into the *PbAP2-G* locus thus repairing the m9 mutation.

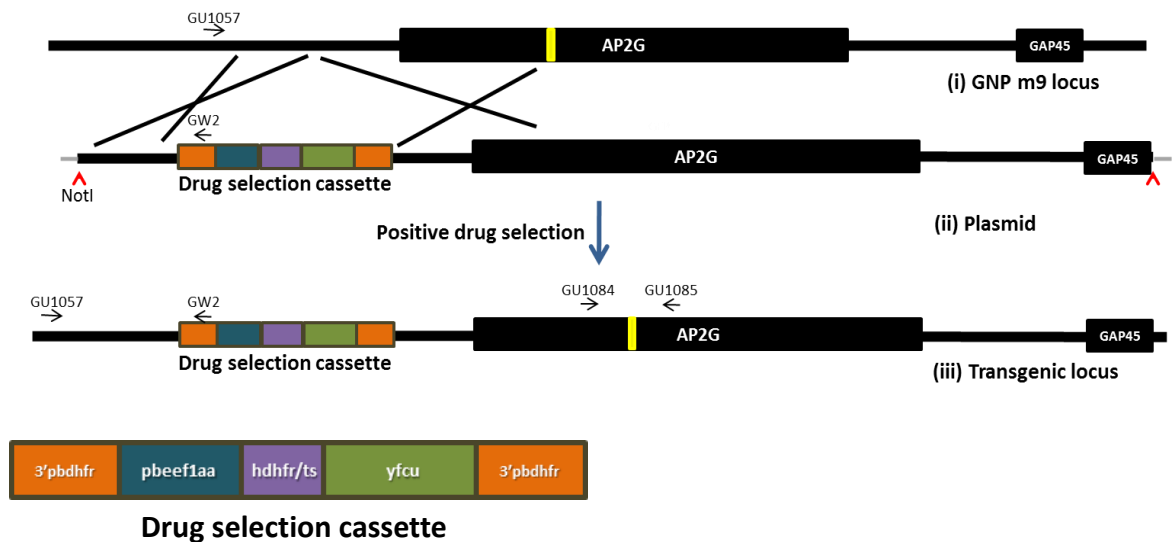
The presence of a few non-specific bands in the gel picture (Figure 6-1b) might be due to low annealing temperature used for the PCRs. But despite the observation of these non-specific bands, the presence of strong integration bands of 1.5 kb for G366-8 and their absence in the GNP m9 and water samples strongly indicate the successful integration of the complementation construct into the desired genomic locus.

#### **6.2.1.1.1      *Phenotypic and genotypic characterization of the PbG01\_COMP-UP complemented natural PbAP2-G GNP mutants***

Despite the PCR-based evidence suggesting a successful 5' integration of the complementation construct (PbG01\_COMP-UP) into the GNP m9 *PbAP2-G* locus, there was no evidence of any gametocytes on Giemsa-stained smears and flow cytometry in the transfectants, G366-8. It was not possible to examine the 3' integration of the construct at the desired genomic locus by PCR-based methods because the very large region of genomic similarity between the construct and the target genomic locus generated a large expected product (~8 kb) that is difficult to PCR-amplify. In the absence of confirmation of the 3' integration, the precise point of integration at the 3' end could not be confirmed. However, the site of an integration event at the 3' end is important as this might determine the success of complementation and the restoration of the WT phenotype. If the site of integration is anywhere upstream to the m9 SNP locus, there will not be any restoration of the phenotype since the transfectants will still retain the m9 SNP despite a successful integration at both ends (Figure 6-2). Conversely, if the integration happens to occur anywhere downstream to the m9 SNP, the restoration to the WT phenotype becomes highly likely (Figure 6-1; see GNP m9 locus (i) and the transgenic locus (ii)). Whatever the case may be, there should not be any other mutations than that in question (m9 SNP) conferring the GNP phenotype as evident in the cloned m9 GNP line.

There is a third possibility of getting a mixed population of parasites some having a proximal and some distal (as compared to the location of the m9 SNP) integration at the 3' end. If this happens, the ultimate phenotype of the transfectants will most likely be that of the gametocyte non-producers, given

the fact that the GNP phenotype offers a growth advantage over the gametocyte producer phenotype. Thus, any population of the GNP type will tend to overgrow the gametocyte producer phenotype population and will eventually replace the latter, thus resulting in the pure GNP phenotype.

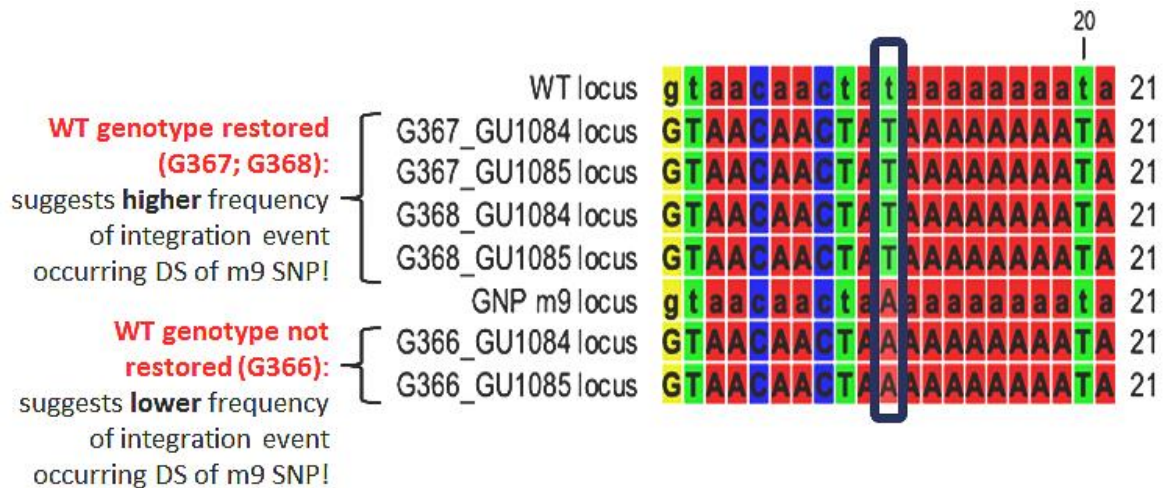


**Figure 6-2. Schematic of complementation strategy used to repair the m9 mutation (yellow bar) in the *PbAP2-G* (not-to-scale) showing mutated endogenous GNP m9 & modified genetic locus together with the recombineering based plasmid construct (PbG01\_COMP-UP) and the drug selectable marker (multi-coloured bar; orange=3'UTR of Pbdhfr; dark blue=pbeef1aa; violet=hdhfr; green=yfcu). The schematic highlights the importance of the site of homologous recombination event at the 3' end. Note that in this case, the recombination event is shown to occur upstream of the m9 mutation and hence the resultant transgenic locus (iii) retains the m9 mutation and hence the GNP phenotype. This is in contrast to the 3' homologous recombination event happening downstream of the m9 mutation (Figure 7.1(a)), resulting in the restoration of the m9 locus to the WT thus reverting the GNP phenotype to the WT phenotype (successful complementation).**

Considering all of the above possibilities, following evidence of successful 5' integration, the way to successfully isolate mutant parasites having the correct version of the 3' integration is to clone the repaired lines first and then identify the correct clone by sequencing through the m9 mutation region. A better alternative (in terms of saving time and animals) is to first sequence the repaired mutants and then clone the best one from the mutants. The latter strategy was chosen and all three transfectants (G366-8) were sequenced. The region (~1000 bp) flanking the m9 SNP (around 650 bp upstream and 350 bp downstream to the m9 SNP) was PCR-amplified using high fidelity Taq polymerase (Expand High Fidelity PCR System; Roche) and the oligonucleotide primers GU1084/GU1085 (Figure 6-3). As evident from Figure 6-3, the m9 SNP locus was restored to the WT genotype in lines G367 and G368 whereas line G366



retained the m9 GNP genotype and phenotype. A possible explanation of this may come from the site of recombination event happening at the 3' end (as explained in section 6.2.1.1.1) suggesting that a higher frequency of the recombination event downstream of the m9 locus favours restoration of the m9 genotype and phenotype to the wild type and vice versa. The lines G367 and G368 were subsequently successfully cloned by the limiting dilution method (please see section 2.2.11 of Materials & Methods) and the clones G367cl2 and G368cl1 were selected for further work on complementing these lines in future.



**Figure 6-3. Multiple sequence alignment using CLC version 5.0 software.** The genomic sequences (cropped from left and right from the original) from the PCR-amplified region encompassing the m9 locus from the lines G366-8 using oligonucleotide primers GU1084/GU1085 were aligned to examine the possible site of 3' recombination event. The black vertical box highlights the nucleotide base involved in the m9 T-A SNP. The WT locus had a thymine which was mutated to adenine in the GNP mutant m9. As evident, the lines G367 and G368 had the mutation corrected whereas G366 retained the mutation suggesting a higher frequency of a 3' integration event downstream of the m9 SNP. In comparison, a higher frequency of a 3' integration event upstream to the m9 SNP locus might have been responsible for the non-restoration of the m9 SNP relative to the WT genotype.

#### 6.2.1.1.2 Possible explanation for failed complementation in G367 and G368

Despite sufficient evidence suggesting the presence of a restoration-favourable genotype in G367 and G368, there was no evidence of successful complementation in terms of restoration of the GNP phenotype to the wild type phenotype in either of the lines. The possible reasons for a failure to complement were examined. A potential interruption of the AP2-G promoter region by the drug-selectable marker cassette (~3 kb) was suspected to be one of the causes. It could also be due to altered length of the endogenous AP2-G

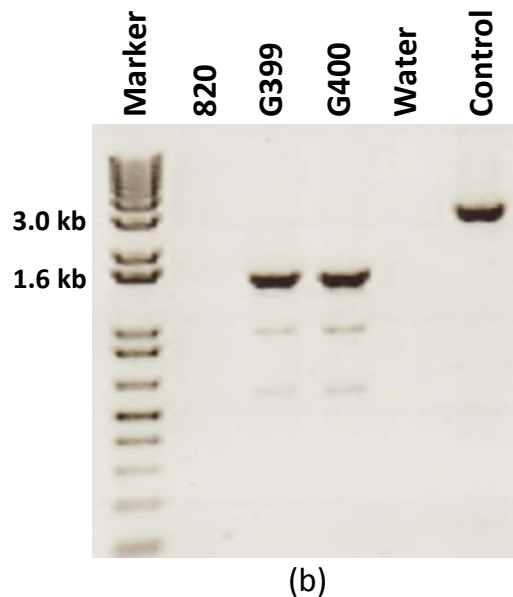
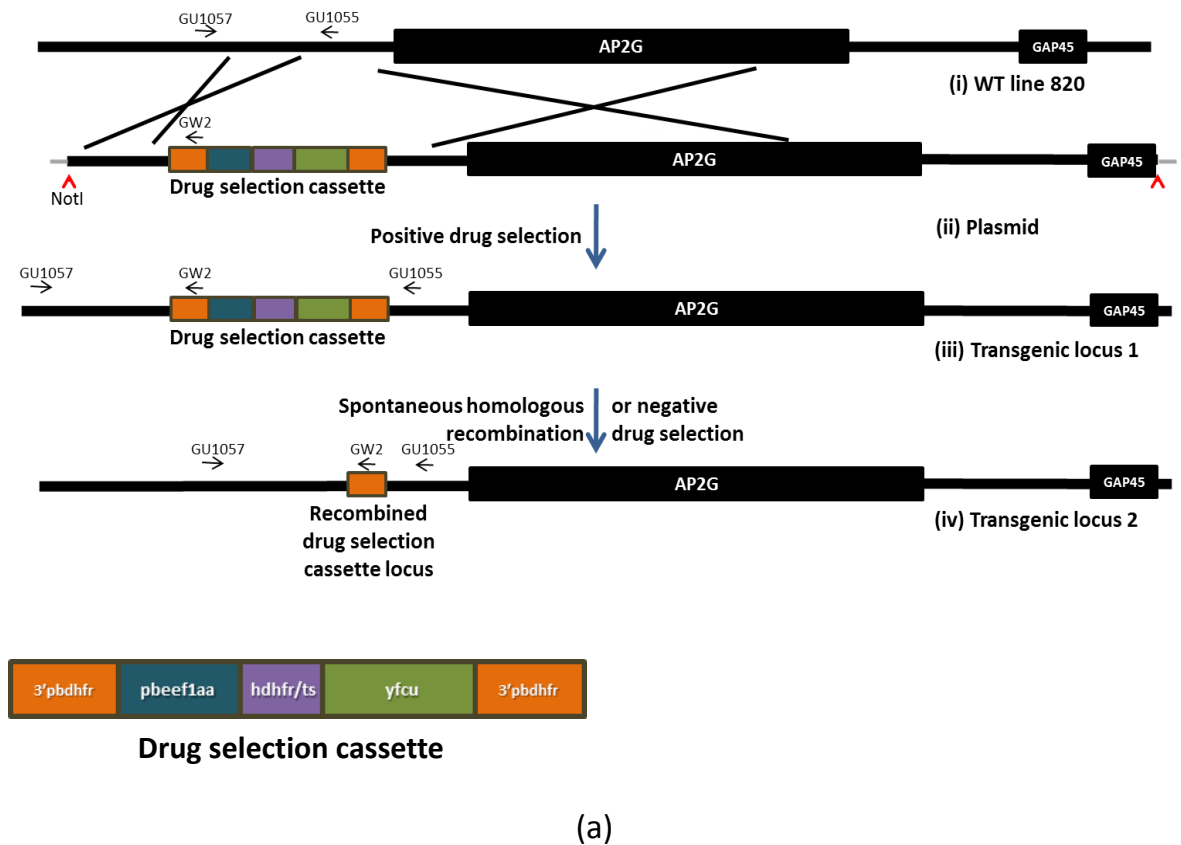
promoter rendering a conformational change in the way the DNA molecules are coiled around histone proteins (nucleosome phasing). It is well known that nucleosomal organization at the genomic level has a critical role in regulating gene expression in eukaryotes (Kornberg and Lorch, 1999). There is a particular ‘phasing’ of the nucleosome in relation to the transcription start site (TSS) at the genomic level with an average inter-nucleosomal distance of 162 bp (Cui et al., 2012). Any alteration in nucleosome phasing may put the nucleosome out of phase in relation to the TSS and has been shown to be directly correlated with gene expression by either increasing or impeding the accessibility of functional *cis*-elements to RNA polymerases II or transcription factors or both (Hasan, S. 2003, PhD thesis; Schones et al., 2008).

To verify the hypothesis that an interruption in the *AP2-G* promoter region might have contributed to failure to complement the GNP m9 line, the parental WT line 820 was transfected with ~5 µg of PbG01\_COMP-UP DNA (Table 6-2; Figure 6-4).

Table 6-2

Oligo set	Details	Expected band size with DNA (kb)				Ta (°C)	Ext Time (min)
		WT 820	G399-G400	Water	Control		
GW2/1057	5'Integration	--	1.5	--	NA	48	5
1233/1242	Control	NA	NA	NA	2.9	48	5

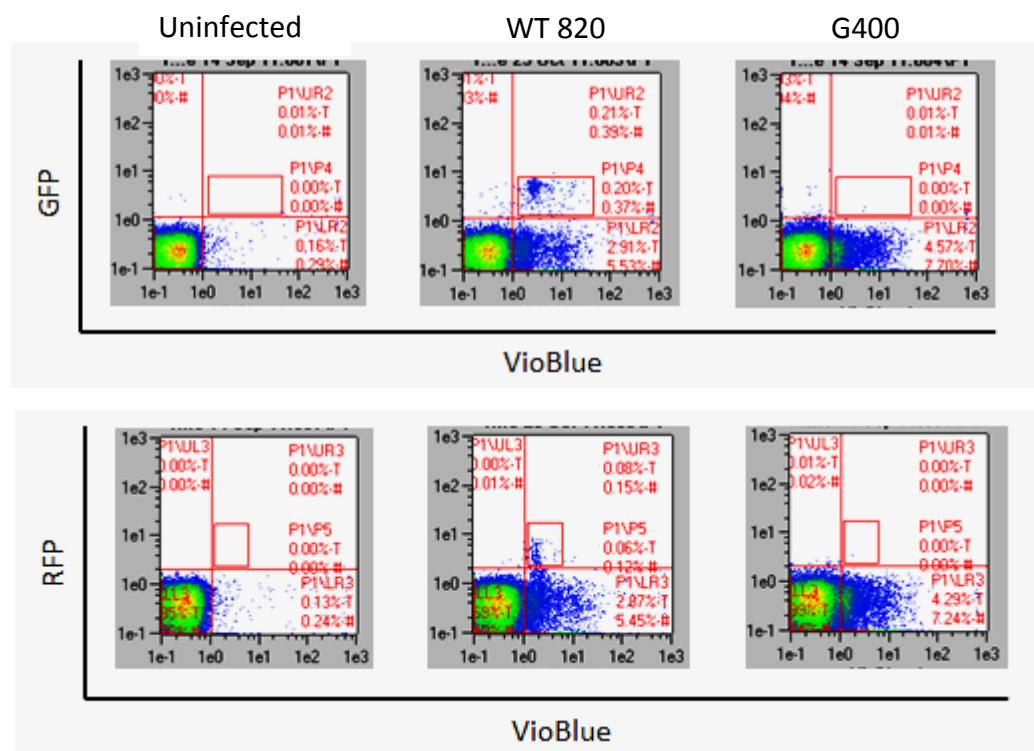
**Table 6-2. Basic PCR reaction settings for amplifying various regions of the *PbAP2-G* genomic locus to identify the presence of 5' integration of the recombineering based complementation construct (PbG01 COMP-UP) into the WT line 820 genome.**



**Figure 6-4 (a) Schematic (not-to-scale) of transfection strategy used to test the role of *PbAP2-G* promoter interruption in the failure of complementation of GNP m9 line with *PbG01\_COMP-UP*.** The wild type line 820 (i) was transfected with the recombineering based plasmid construct *PbG01\_COMP-UP* (ii) containing the drug selectable marker cassette (multi-coloured bar; orange=3'UTR of *Pbdhfr*; dark blue=*pbeef1aa*; blue=*hdhfr*; green=*yfcu*) upstream to *AP2-G*. The resultant positively selected mutant (iii) is similar to the WT except for the presence of the drug selectable marker cassette sitting 1.2 kb upstream to the *AP2-G* TSS. The schematic highlights the interruption of the *PbAP2-G* promoter with the selection cassette. **Figure 7.4(b) Diagnostic PCR.** The PCR confirmed successful 5' integration using the following primers GU1057/GW2 (5' integration band = 1.5 kb). Strong positive bands of desired size were observed for both the mutants indicating the integration of the construct into the *PbAP2-G* locus thus interrupting the promoter of *PbAP2-G*.

### 6.2.1.1.3 Genotypic and phenotypic confirmation of G399 and G400

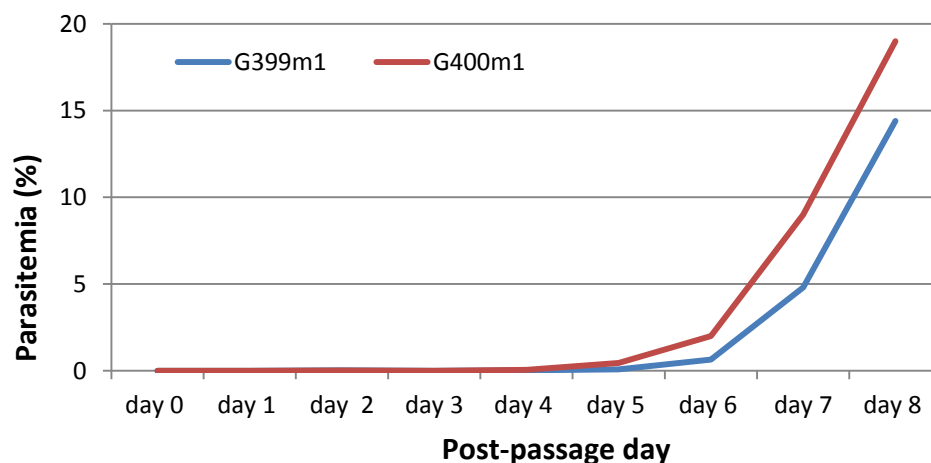
After positive selection with oral pyrimethamine, two transfectants (G399-G400) were obtained; both of them were screened by PCR (Table 6-2 and Figure 6-4b) which was suggestive of successful 5' integration of the construct into the WT genome thus interrupting the promoter. Phenotypically, both G399 and G400 had no observable gametocytes on Giemsa-stained peripheral blood smears and flow cytometry (Figure 6-5) and thus had the classical GNP phenotype, similar to that observed for the *PbAP2-G* knockouts (See Chapter 5). Although the precise length of the promoters in *Plasmodium* are not known, so as to put the position of the SM cassette in context of the probable promoter region, an upstream sequence length of 2 kb is conventionally used as a standard promoter (Mair et al., 2009; Ponzi et al., 2010). Other conditions being similar, the resulting GNP phenotype in G399 and G400 strongly supports the assertion that indeed the promoter of the *AP2-G* was interrupted by the SM cassette while attempting complementation with PbG01\_COMP-UP construct.



**Figure 6-5. FACS plots showing the GFP signals (Y-axis; upper panel; male gametocytes) and RFP signals (Y-axis; lower panel; female gametocytes) detected in FITC and PEA channels, respectively in the Hoechst-positive population (X-axis) in the blood from uninfected, WT line 820 infected and the G400 line. The G400 line (and G399; not shown) showed no gametocytes on flow cytometry as compared to the WT line 820 thus suggesting that the G399 and G400 lines were GNP lines.**

#### 6.2.1.1.4 Repairing the promoter in lines G399 and G400

Various attempts to repair the promoter interruption in G399 and G400 were tried including a negative selection experiment to reduce the length of the interrupting SM cassette thus providing a chance (although bleak) of putting the nucleosomes back in phase with AP2-G TSS. The fact that the selectable marker cassette had two identical copies of the 3'*pbdhfr/ts* flanking the *pbeef1aa*, *hdhfr/ts* and *yfcu*, negative selection with 5-FC (please see section 2.1.5 of Materials & Methods) was attempted to recycle the selectable marker cassette thus leaving a footprint of only ~450 bp into the genome (Figure 6-4 (iv)). 5-FC was used from day0 post-passaging (Figure 6-6). Following negative selection on G399 and G400, the two lines obtained were labelled G399m1 and G400m1. The parasitemia kinetics of both G399m1 and G400m1 in Figure 6-6 suggests that either the majority of the pre-negative selection parasite population already had the selectable marker cassette recycled, possibly because the recombination event happened spontaneously, or there was a minimal WT contamination in G399 and G400 that was high enough to reflect gametocytemia in FACS (Figure 6-5) as G399 and G400 were not cloned before proceeding for negative selection. The decision of proceeding with the uncloned lines was based on the FACS observation of a GNP phenotype and on the assumption that GNP phenotype will eventually overgrow the WT counterpart, if any, over time thus self-cloning the GNP parasites.



**Figure 6-6. Negative selection for lines G399 and G400 using 1.5 mg/ml 5-FC dissolved in drinking water.** The X-axis shows days after passage of the respective lines; 5-FC was administered from the day of the passage in drinking water and continued till day 8. Y-axis shows the parasitemia as percentage of infected red blood cells. The steep rise in parasitemia in absence of an initial peak suggested that the majority of the starting population had already undergone spontaneous homologous recombination event between the two identical copies of 3'*Pbdhfr/ts* before 5-FC selection.

PCR analyses of the negatively selected lines G399m1 and G400m1 suggested the latter possibility of the presence of a WT population which somehow “escaped” from being overgrown by the GNP counterparts and became selected (over lines with the full drug selectable marker or with the ~450 bp post-negative selection footprint) during negative selection (Table 6-3; Figure 6-7). The FACS analyses showed gametocytes both in G399m1 (Figure not shown) and G400m1 (Figure 6-8) suggesting success in repairing the interrupted *PbAP2-G* promoter by reducing the size of the interruption. But the evidence from PCR analyses supported the amplification of the WT genotype instead producing the gametocytes. The possibility of both scenarios is intriguing - assuming the observation of gametocytes in flow cytometry was due to the selection and subsequent proliferation of the latent WT population, the absence of any reflection of the WT contamination in the FACS data from G399 and G400 argues against this assumption. This would also mean that the GNP mutants failed to overgrow the WT counterpart in G399 and G400 which cannot be explained. Thirdly, it also implies that it is possible to select for a latent gametocyte producing WT population even against the definite growth advantage of the dominant GNP population. If this was the case, then it would seem possible to use negative selection for selective proliferation of a complemented (WT-like) gametocyte producer population against a GNP background, which would provide an advantageous determinant of success of complementation (using negative selection) in the engineered GNP mutants.

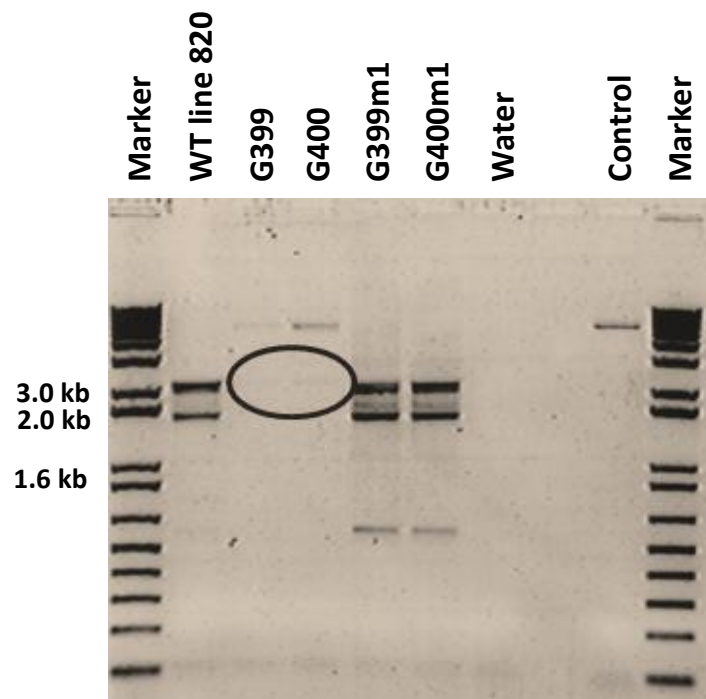
Table 6-3

Oligo set	DNA	Expected band size (kb)	Ta (°C)	Ext Time (Min)
1055/1057	WT line 820	2.1	54	7
1055/1057	G399	5.5	54	7
1055/1057	G400	5.5	54	7
1055/1057	G399m1	2.7	54	7
1055/1057	G400m1	2.7	54	7
1055/1057	Water	--	54	7
1233/1235	Control	5.9	54	7

**Table 6-3. Basic PCR reaction settings** for amplifying the region across selectable marker cassette positioned in the immediate 5' upstream region of the *PbAP2-G* genomic locus in lines WT 820, G399 and G400 and their negatively selected counterparts, G399m1 and G400m1 to identify the efficacy of the negative selection in terms of a reduction of the size of the promoter interruption.

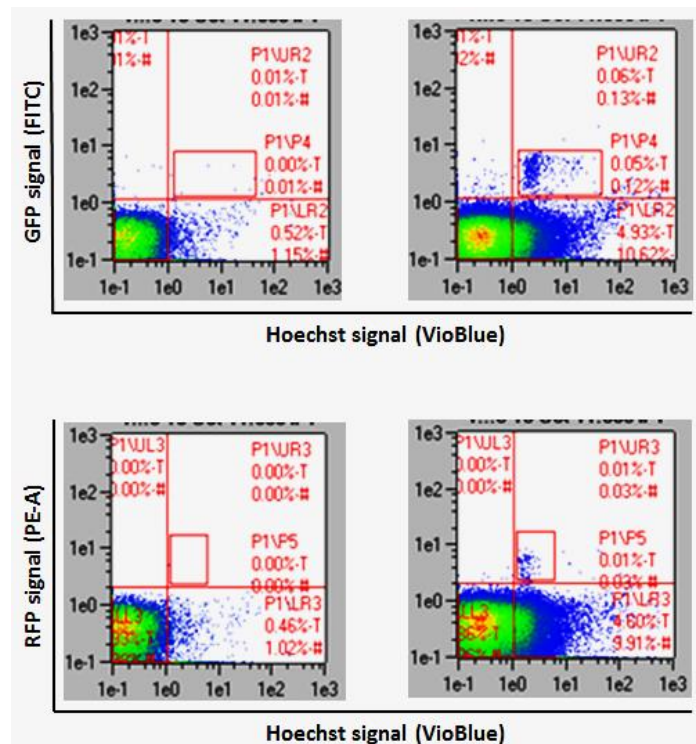
### 6.2.1.1.5 Negative selection of cloned repaired lines to reduce the *PbAP2-G* promoter interruption

Because of the uncertainty of success involved in the repair (reducing the size) of the interruption in the *PbAP2-G* promoter, the cloned repaired lines (G367cl2 and G368cl1) were subjected to negative selection. This generated the lines G367cl2m1 and G368cl1m1 which when examined under Giemsa-stained peripheral blood smears and flow cytometry showed no evidence of gametocytemia and hence no evidence of successful complementation (data not shown). Although the negative selection successfully reduced the size of the interruption of the *PbAP2-G* promoter region, even the presence of ~450 bp of footprint was sufficient to prevent the expression of *AP2-G* and hence the success of complementation of natural GNP mutants with the PbG01\_COMP-UP construct.



**Figure 6-7. Diagnostic PCR across the position of the selectable marker cassette in the immediate 5' upstream region of the *PbAP2-G* genomic locus in lines WT 820, G399, G400 and their negatively selected counterparts, G399m1 and G400m1 to identify the efficacy of the negative selection in terms of a reduction of the size of the promoter interruption.** The lines G399 and G400 show minimal contamination of the WT band (2.1 kb) as these were still uncloned lines together with the band size corresponding to the full length selectable marker cassette (5.5 kb). The negatively selected lines G399m1 and G400m1 instead of showing a predominant 2.7 kb band suggesting marker recycling (thus reducing the size of the selectable marker interruption) showed a stronger WT band (2.1 kb) suggesting that the WT residual in G399 and G400 selectively proliferated under negative selection in G399m1 and G400m1 rather than the selectable marker recycling, thus hampering the possibility of repair of the interrupted promoter in lines G399m1 and G400m1.



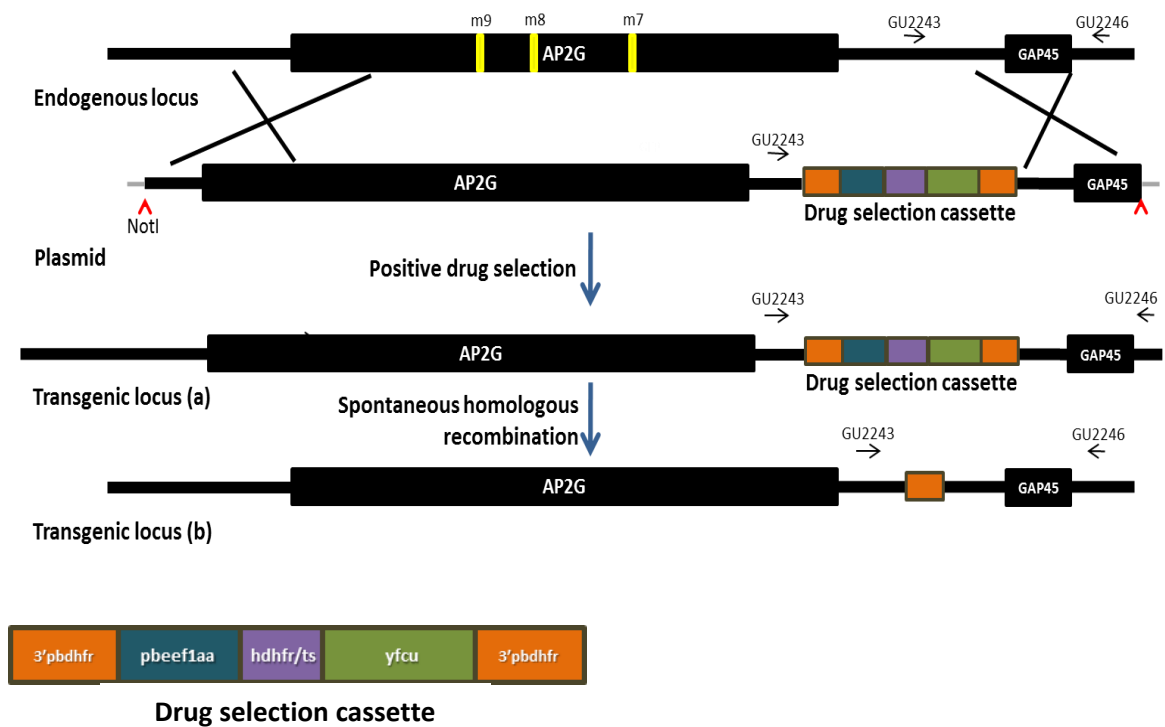


**Figure 6-8. Flowcytometry plots showing presence of gametocytes examined in blood from line G400m1 (right pane) as compared to the blood from uninfected mouse (left pane).** The X-axis shows the Hoechst signals in the VioBlue channel in both upper and lower panels whereas the vertical axis in the upper panel shows the GFP signals (male gametocytes) in the FITC channel and RFP signals (female gametocytes) in the PE-A channel. The presence of male and female gametocytes in the line G400m1 suggests the repair of the promoter and thus restored functioning of *PbAP2-G* although the reactivation of the residual WT parasites from line G400 in line G400m1 was suspected of generating FACS-detectable gametocytemia in an apparently repaired line.

### 6.2.1.2 Complementation of natural *PbAP2-G* GNP mutants with PbG01\_COMP-DOWN

Looking collectively at the body of evidence regarding successful integration of the huge recombineering-based complementation construct and regarding the possible interruption of the *PbAP2-G* promoter region with the PbG01\_COMP-UP construct which might have hampered the success of the complementation attempts, it was felt essential that a new complementation construct be made such that the drug selectable marker cassette sits within the 3'UTR of *PbAP2-G* instead of the promoter region. This new construct was again kindly developed and gifted to me by Oliver Billker and his group at Sanger. This new construct was called PbG01\_COMP-DOWN and had the drug selectable marker cassette sitting ~2 kb downstream of the *PbAP2-G* stop codon (Figure 6-9).





**Figure 6-9. Schematic (not-to-scale) of the complementation strategy used to repair the naturally acquired GNP mutations in lines m9, m8 and m7.** The WT endogenous locus shows full length *PbAP2-G* with relative location of the three natural GNP mutations (yellow vertical bars). Also shown below is the NotI digestable complementation construct (PbG01\_COMP-DOWN) bearing the full length WT copy of *PbAP2-G* with a selectable marker cassette (multi-coloured horizontal bar; detailed at the bottom right corner of the figure; orange=3'UTR of *Pbdhfr*; dark blue=pbeef1aa; violet=hdhfr; green=yfcu) positioned approximately 2 kb downstream to the *AP2-G* stop codon. The recombinable homology regions between the endogenous locus and the construct are depicted as crosses. The recombined transgenic locus (a) selected using pyrimethamine as positive selection shows the repaired mutations in the genome together with the introduction of the selectable marker cassette downstream. The modified transgenic locus (b) shows the occurrence of possible spontaneous recombination event between two homologous copies of the 3' *Pbdhfr*/ts resulting in elimination of the part of the cassette flanked by the two copies thus reducing the size of the interruption. The position of the oligonucleotides used for checking the integration of the construct into the genome is also shown (Table 6-4).

Transfections were done with ~5 µg of NotI (New England BioLabs) linearized PbG01\_COMP-DOWN constructs in triplicate into WT line 820 (G657; as a control to examine if the positioning of the selectable marker cassette had any effects on the expression of *PbAP2-G*) and each of the three natural *de novo* GNP lines; GNP line 9 (G654, G655 & G656), GNP line 8 (G699, G702 & G703) and GNP line 7 (G696, G697 & G698). The resulting transfectants were selected by using pyrimethamine in the drinking water. Lines G656 and G657 were subsequently cloned and clones G656cl1-2, and G657cl1-3 together with transfectant lines derived from GNP lines 7 and 8 PCR-screened (for successful integration of the complementation construct into the genome by amplifying the region spanning through the selectable marker cassette using oligonucleotide primer pairs

G2243/GU2246 as shown in the Figure 6-9. The expected length of the amplicons for each of the integration events was 5.5 kb as compared to 2.0 kb each in the WT line 820 and the parental GNP lines 9, 8 and 7 (Table 6-4). A summary of all the oligonucleotide sets used as primers in the PCRs, their details, expected length of bands, annealing temperature (Ta) settings used and extension time used for various PCR reactions are shown in Table 6-4.

Table 6-4

Oligo set	Details	Expected band size with DNA (kb)					Ta (°C)	Ext Time (min)
		WT line 820	GNP	Transfectants	Water	HP		
2243/2246	Through the selectable marker	2.0	2.0	5.5	--	NA	52	6
1233/1242	Control	NA	NA	NA	NA	2.9	52	3

**Table 6-4. Basic PCR reaction settings for amplifying the region across the selectable marker cassette in the endogenous locus (WT line 820 and GNP lines) and transgenic loci (a) and (b) (complemented lines; Figure 6-9).** The relative size of the PCR product at the desired locus indicates the success of integration of the construct into the targeted genomic locus.

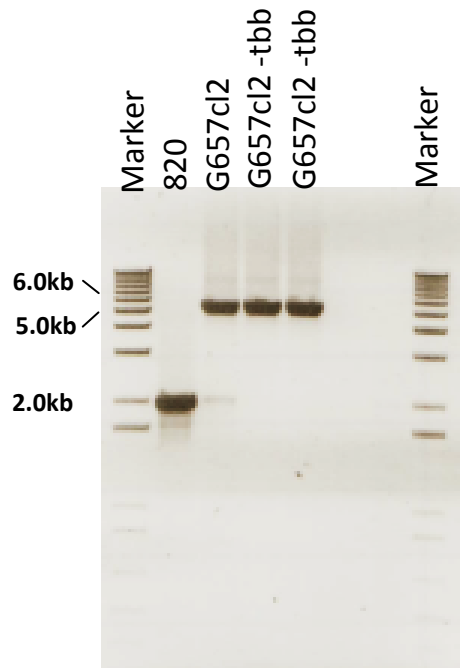
#### **6.2.1.2.1 Genotypic and phenotypic characterization of the GNP mutants complemented with PbG01\_COMP-DOWN**

PCR analyses through the integration site using oligonucleotide primers GU2243/GU2246 from the lines WT 820, natural GNPs and their corresponding repaired lines showed expected results. As discussed earlier, the reasons for the failure to complement natural GNP mutants with the upstream construct PbG01\_COMP-UP and hence the need for the new downstream construct (PbG01\_COMP-DOWN) were at least two-fold - first, the correction of the GNP mutation was dependent on the precise site for the 3' integration of the complementation construct into the genome, and second, more importantly, any successful genotypic correction of the GNP mutation was masked by the inability of *PbAP2-G* expression stemming from the interruption of the promoter region of the gene by the drug selectable marker cassette built into the plasmid. The downstream version of the complementation construct PbG01\_COMP-DOWN was

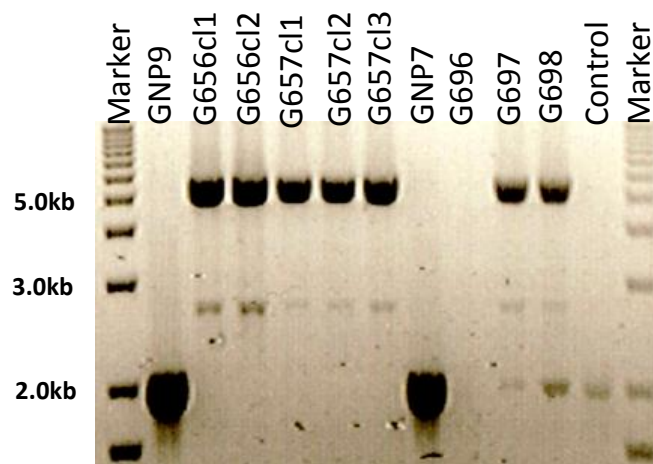
designed to solve the second problem (by maintaining the integrity of the *PbAP2-G* promoter) so that any corrective homologous recombination event results in expression of the repaired gene copy. This possibility of non-interference of the expression of the locus by the complementation construct was examined by transfecting the construct into the WT line 820. The hypothesis being, any interference of expression due to the position of the SM cassette will be reflected in the transfectants' GNP phenotype. If the number of gametocytes formed in these transfectants is optimum, it would imply that the position of the SM cassette is not interfering with the gene expression and thus the construct could be utilized to functionally repair the GNP mutation.

The PCR analyses from the three clones of line G657 revealed that the construct integrated into the genome as evidenced by the presence of the 5.5 kb band in Figure 6-10(a). There was no WT background band in all three clones, compared to the WT band of 2.0 kb observed for the GNP9 line (which is similar to the WT 820 line for the region amplified and hence used as a reference here). There is however the presence of a smaller less predominant ~2.7 kb band for all the three clones; this might be a reflection of a spontaneous homologous recombination event occurring between the two identical copies of the 3'UTR of *Pbdhfr/ts* present in the construct.

Once confirmed that the complementation cassette had successfully integrated into the genome, the phenotype of the three clones was examined using Giemsa-stained smears and flow cytometry (Figure 6-11). G567 clones, except G567cl1 revealed optimal gametocyte production which suggests that there is no hindrance to gene expression in the presence of the SM cassette within the 3'UTR of *PbAP2-G*. The three natural GNP lines that were transfected with the PbG01\_COMP-DOWN construct (G655cl1 and G656cl1 from GNP9 & G696, G697 and G698 from GNP7) showed a similar genotype on PCR analyses, except for G696 (Figure 6-10b; G654 and G702, G703 and G704 are not shown).

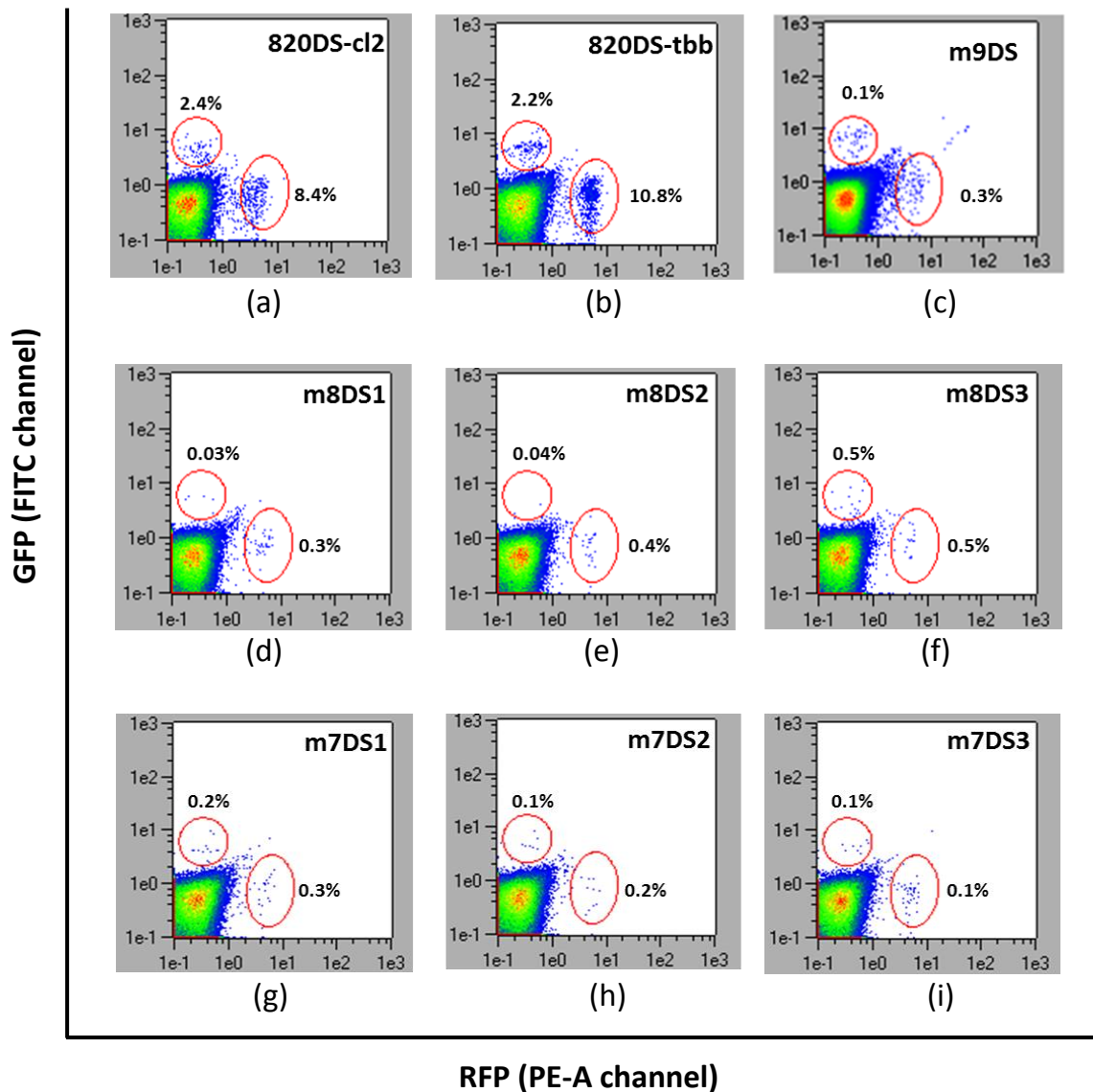


(a)



(b)

**Figure 6-10 (a-b). PCR analyses of the line WT 820 and lines WT 820, GNP9 and GNP7 transfected with PbG01\_COMP-DOWN.** The region of the genome spanning the locus of the selectable marker cassette was amplified in each of these lines. (a) shows the corresponding PCR-amplified bands in line WT 820, G657 cloned and G657 transmission bite-back (tbb). Band in the WT line 820 (2.0 kb) suggests no evidence of the selectable cassette whereas the bands in the transgenic line G657cl2 (820DS) and the G657cl2tbb lines (820DS-tbb) show the modified locus corresponding to the size of the entire cassette (5.7 kb) which strongly suggested successful integration and optimum gametocyte production by these lines. In figure (b), bands corresponding to the WT line size (2.0 kb) were observed in the GNP lines 9 and 7. The 5.7 kb bands suggesting successful integration could be observed for all the transgenic lines examined except G696. A fainter 2.7 kb band was also observed in all the transgenic lines except G696 suggesting spontaneous recycling of the selectable marker cassette. The transgenic lines G697 and G698 also show the band corresponding to the WT size as these lines were still uncloned and hence had some WT contamination. Overall, the evidence is highly suggestive of successful integration of the construct in all the transgenic lines.



**Figure 6-11. Flow cytometry plots showing gametocytemia levels in the lines transfected with the complementation construct PbG01\_COMP-DOWN.** On the horizontal axis are the RFP signals (female gametocytes) as detected in the PE-A channel and on the vertical axis are the GFP signals (male gametocytes) as detected in the FITC channel. Quite obvious is the optimum number of gametocytes generated by the “complemented” WT 820 line G657cl2 (labelled 820DS; Fig 6-11a), the functional sufficiency of which was evidenced by the number of gametocytes observed when the same line was transmitted through mosquitoes generating line G657cl2-tbb (labelled 820DS-tbb; Fig 7-11b). Taken all together, this shows that the construct PbG01\_COMP-DOWN is functionally capable of restoring the phenotype without affecting the expression of the gene, *PbAP2-G*. A sub-optimal number of gametocytes was observed in the natural GNP lines 9, 8, and 7 when transfected with the same construct generating lines G656cl2 (labelled m9DS), G702-4 (labelled m8DS1-3) and lines G696-9 (labelled m7DS1-3).

The phenotypic analyses of the genotypically “repaired” lines by Giemsa-stained peripheral blood smears and flow cytometry however showed very few gametocytes as compared to the WT line 820 “repaired” using the same construct as a control (Figure 6-11 and 6-12). Furthermore, the number of

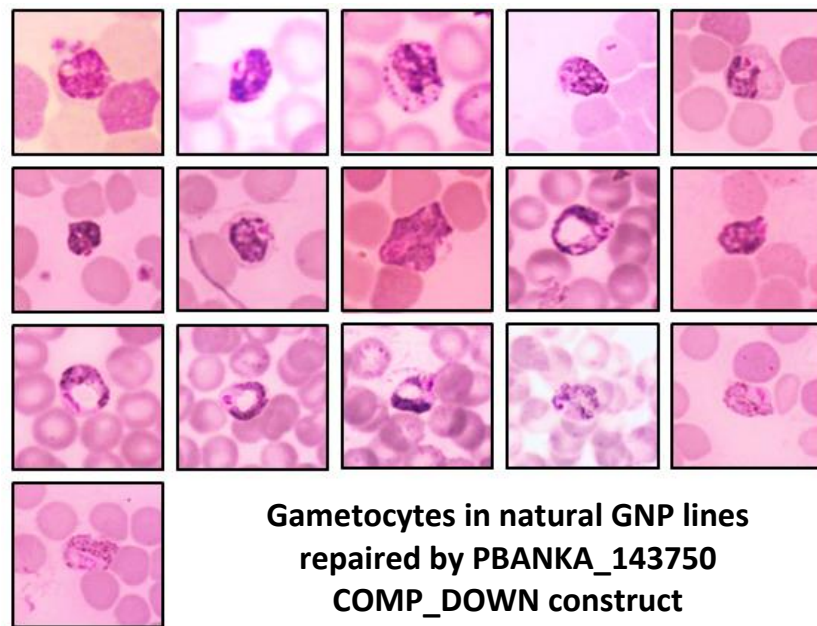
gametocytes got lower and lower as these repaired GNP lines were continued to be grown or passaged (data not shown). This was highly suggestive of the possibility of the presence of a mix of two populations with differential asexual growth potential which could result from such a transfection - one which does not produce gametocytes (derived from the persistence of the parental GNP population as these repaired lines were not yet cloned) and the other which does produce gametocytes with unknown potential (derived from the genotypically repaired GNP population). As discussed earlier, the potential of the GNP population to quickly overgrow any gametocyte producer population and eventually replace them over time could be the main factor responsible for the observed sub-optimal and declining gametocyte producer phenotype.

One immediate solution to rescue the second favourable type of population (the gametocyte producer population) from the mixed population was to either immediately clone the lines without further passage (to minimize the risk of losing the gametocyte producing population) or immediately passage the lines through mosquitoes and perform a bite-back experiment which will eventually allow only the gametocyte producers to develop inside the mosquitoes thus rescuing the desired population. Both strategies were attempted - the transfectants were re-activated from their stablate stocks and grown under pyrimethamine (to keep the selection pressure on till cloning). The lines G656 and G657 were cloned by the limiting dilution method (generating lines G656cl1, G656cl2, G657cl1, G657cl2, and G657cl3) while reserving the originally infected mice for mosquito transmission experiments.

Unfortunately, none of the “repaired” GNP lines could be rescued following mosquito passage signifying overgrowth of the gametocyte producer population by the non-producers during the reactivation and propagation of the “repaired” lines resulting in the total loss of gametocyte producers even before the attempted cloning and mosquito transmission experiments.

Taken all this evidence into consideration, repeat experiments were planned and conducted by one of my colleagues in the lab. Basically, the transfections were done again (as above) but the transfectants, as soon as they showed parasites, were immediately cloned by limiting dilution method and also

passed through mosquitoes, without waiting for genotypic confirmation. The mosquito passage offered a good and powerful screen in this case as there was no competing population (other than the desired) that could have transmitted through the mosquitoes. Parasites were observed in Giemsa-stained smears in the bite-backs which demonstrated retrospectively the presence of gametocytes generated by the transfectants and hence successful complementation. The presence of gametocytes was also observed in Giemsa-stained peripheral blood smears and on flow cytometry (Rachael Orr, Andy Waters' group; personal communication).



**Figure 6-12. Giemsa-stained peripheral blood smears from mice infected with the natural GNP lines successfully complemented with the PbG01\_COMP-DOWN construct.** As discussed in the text, all these lines showed very few gametocytes (mostly degenerating gametocytes) as compared with the “complemented” WT line 820 (not shown).

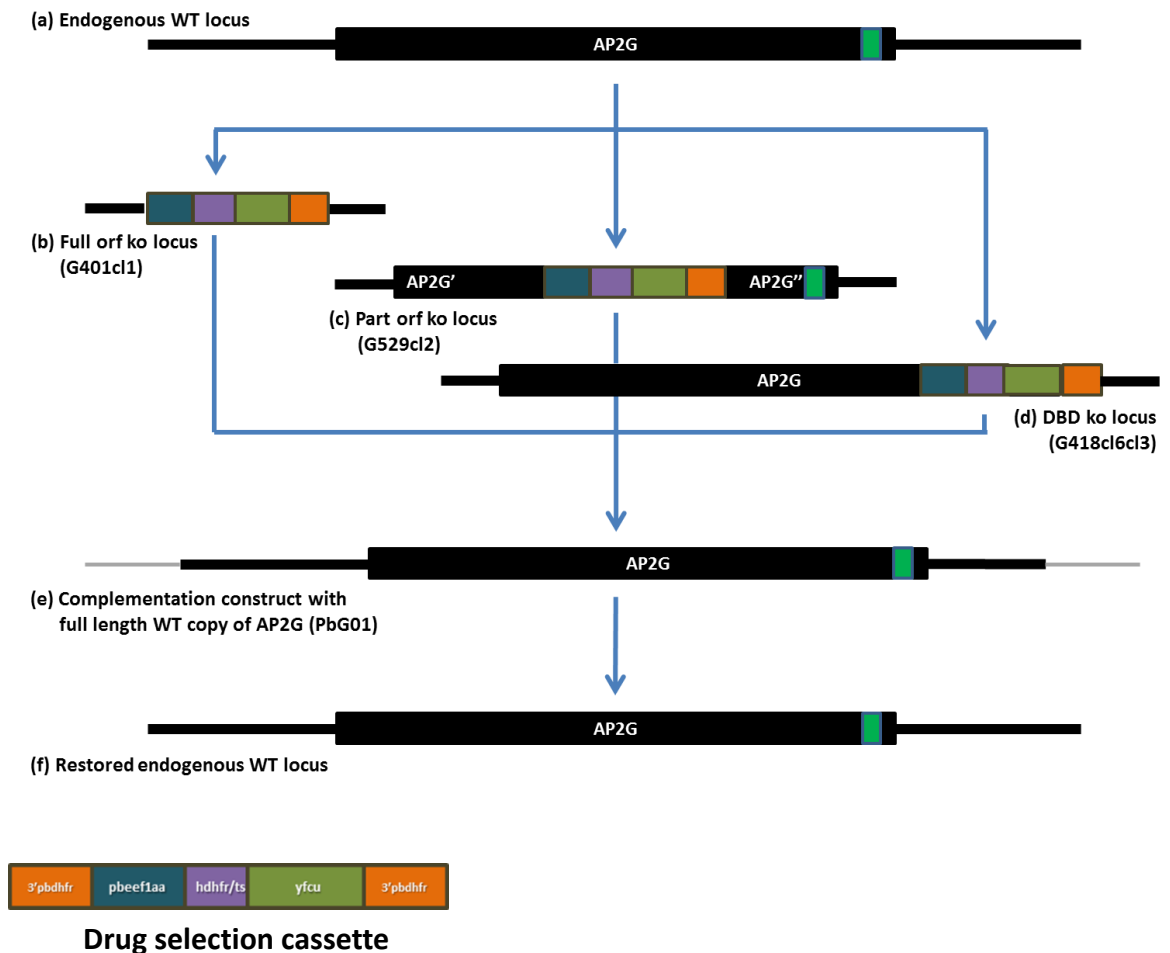
### **6.2.2      *Complementation strategies for genetically engineered PbAP2-G knockout mutants***

Chapter 5 outlined the strategies for knocking the *PbAP2-G* (PBANKA\_143750) gene out which resulted in three cloned knockout length variants of *PbAP2-G* - the full length orf knockout (G401cl1), the partial knockout of the region of the *AP2-G* orf that bears the 3 mutations identified for the naturally selected GNP lines (G529cl2) and the DNA binding domain (DBD) knockout (G418cl6cl3). Complementation strategies involved in repairing these

knockouts differ from that involving the natural GNP mutants in the presence or absence of the positive-negative drug selectable marker cassette (SMC) in the complementation construct. Because the natural GNP mutants did not have the drug SMC in their genome, complementation of these mutants required the presence of drug SMC in the construct to positively select for the complemented parasites (using pyrimethamine) possessing the SMC and hence resistance to pyrimethamine. In contrast, the knockouts already possessed the drug SMC in the genome hence the complementation strategy in these mutants was based on transfecting the construct having a WT functional copy of the whole or part of the orf to be repaired without the drug SMC. The selection of complemented parasites in this case involved the use of 5-FC for targeted killing of the parasites possessing the drug SMC thus conferring sensitivity to 5-FC. However, it should be emphasized here that the drug SMC used for 5-FC mediated target killing must not have two identical copies of the 3'*Pbdhfr*/ts else, instead of 5-FC mediated killing, use of the drug could also result in a homologous recombination event under drug pressure to remove the portion of the drug SMC flanked by identical copies of the 3'*Pbdhfr*/ts thus eliminating the *yfcu* gene which was responsible for conferring the sensitivity to the toxic drug 5-FC and hence failure of the negative selection process and ultimately failed complementation.

Based on the negative selection-mediated complementation of the *PbAP2-G* knockout mutants, two strategies to repair the knockout locus were used - the first involving a recombineering based construct possessing the full-length functional copy of the gene and the second involving a simple mutation-free PCR-product of the part of the orf that needed the repair. Whereas the first strategy could be used to repair all three length variant *PbAP2-G* knockouts, the second strategy could only be used to repair the partial-length *PbAP2-G* knockout and the DBD knockout. The reason why the full length *PbAP2-G* could not be repaired using the PCR-based strategy lies in the length of the gene to be repaired - in this case it was 7.02 kb and it is highly unlikely to obtain such a large PCR product of ~9 kb; 7.02 kb orf plus 1 kb homology arms on each side (Figure 6-13).





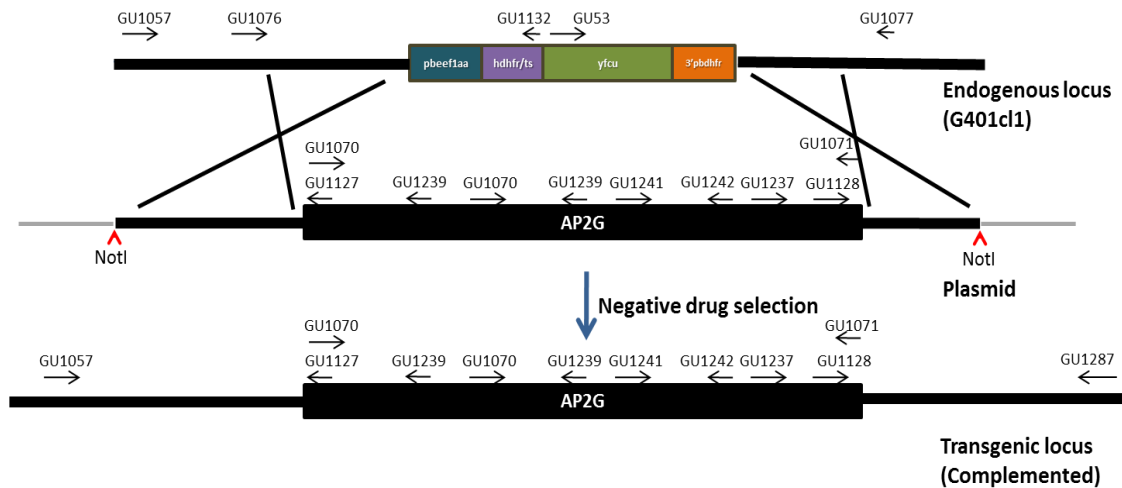
**Figure 6-13. Schematic of complementation strategy used to repair the engineered mutations in the *PbAP2-G* (not-to-scale).** The figure shows the WT endogenous locus (a), genomic loci of the three length variant *AP2-G* knockouts (b to d), the complementation construct used (e; PbG01) and the restored endogenous locus (f). The position of the selectable marker cassette (multi-coloured bar; orange=3'UTR of *Pbdhfr*; dark blue=*pbeef1aa*; violet=*hdhfr*; green=*yfcu*) is shown with each knockout variant and zoomed in for details at the bottom right corner. The DNA Binding Domain (DBD) is shown as a green vertical bar on the *PbAP2-G* orf. The recombineering based complementation construct which was linearized with *NotI* before transfection. The resulting transgenic mutants were negatively selected using the drug 5-FC.

### 6.2.2.1 Complementation of the knockouts using a full length functional copy of *PbAP2-G*

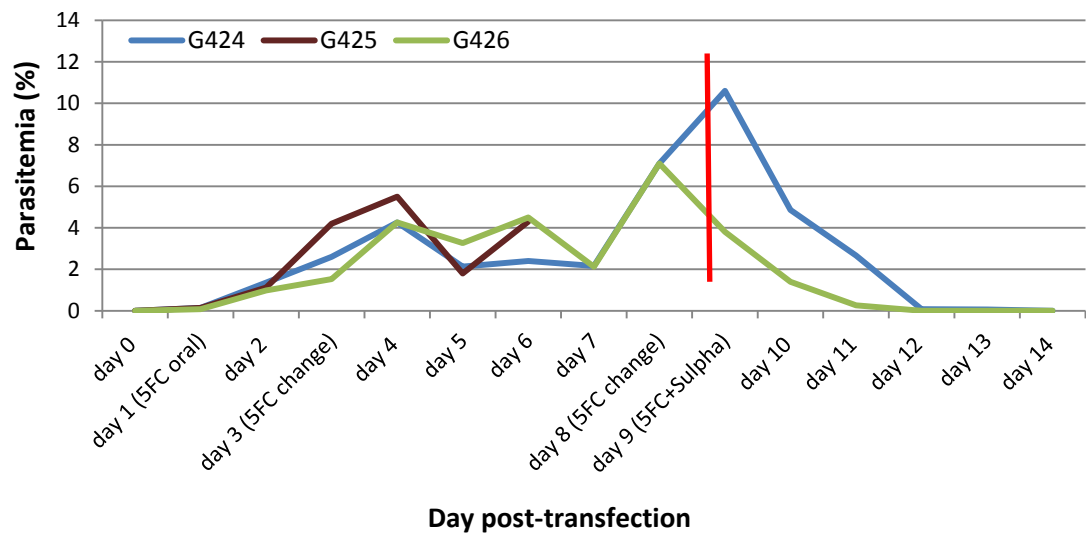
All three length variants of the *PbAP2-G* knockout - G401cl1 (full length knockout), G418cl6cl3 (DBD knockout) and G529cl2 (partial knockout) were subjected to repair using the PbG01 construct.

### 6.2.2.1.1 Complementation in full length *PbAP2-G* knockout *G401cl1*

The complementation construct, called PbG01, bearing the full length wild type functional copy of *PbAP2-G* was generated using recombineering-based approach and kindly supplied by Oliver Billker's Lab (Wellcome Trust Sanger Institute, Cambridge, UK). The construct had full length (7.02 kb) of the WT orf plus upstream and downstream homology arms of ~2.0 kb and ~2.5 kb, respectively (Figure 6-14). In the first round of transfections, approximately 3-5 µg of NotI digested (New England BioLabs) PbG01 was transfected to *G401cl1* schizonts in triplicate and the transfectants (G424, G425 and G426) were negatively selected from day 1 post-transfection (Figure 6-15).



**Figure 6-14. Schematic (not-to-scale) showing the complementation strategy to repair the full length knockout of *PbAP2-G* using full length WT copy of the gene as a recombineering based construct.** The endogenous locus for the knockout (*G401cl1*) had a selectable marker cassette (orange=3'UTR of *Pbdhfr*; dark blue=*pbef1aa*; violet=*hdhfr*; green=*yfcu*) with only one copy of the 3' *Pbdhfr*/ts UTR making the cassette non-recyclable and hence usable for the negative selection-based repair (see text). The vector (PbG01) with two NotI linearization sites is shown. The likely areas for homologous recombination are shown as crosses. The transfectants were negatively selected using the drug 5-FC in drinking water. The resultant complemented transgenic locus is also depicted. Also shown are the relative position of various oligonucleotide primers used for confirmation of integration event (Table 6-5) and the resulting repair.



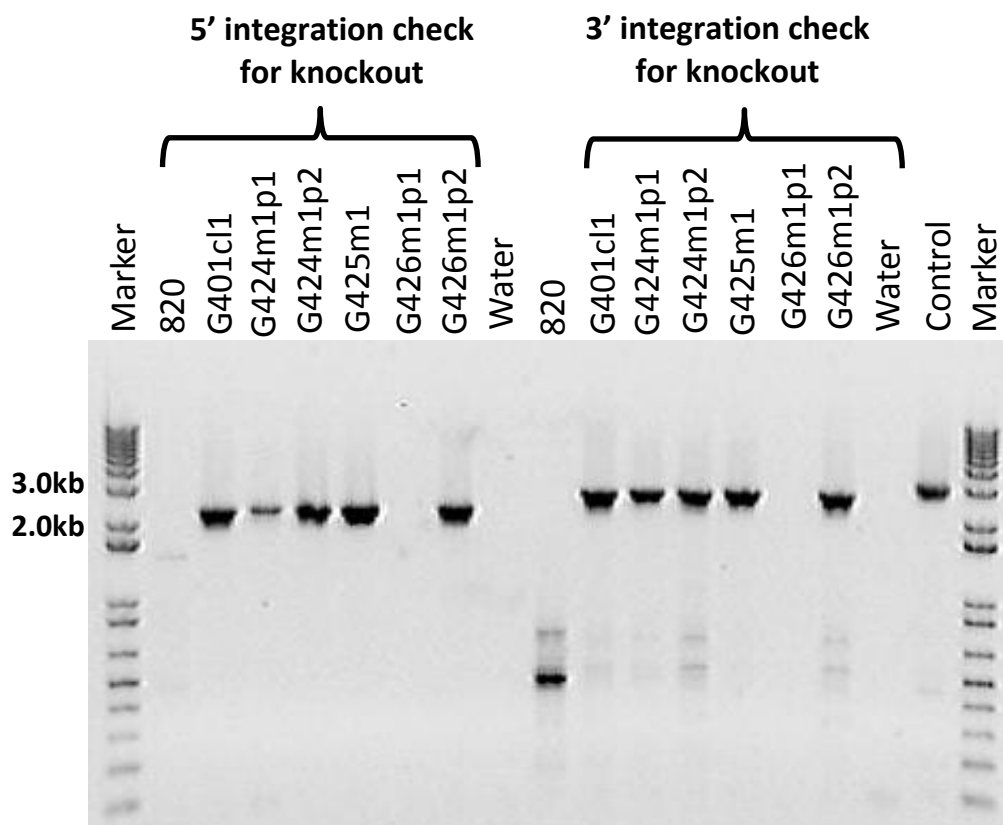
**Figure 6-15. Negative selection experiment on first round of transfectants with PbG01.** The drug 5-FC was started on the day following transfection and was replaced with fresh drug at least every 4<sup>th</sup> day. The X-axis shows the progression of the experiment as days post-transfection and the y-axis shows the total parasitemia, as observed with the Giemsa-stained smears. The vertical red bar denotes the day when the lines were put on Sulphadiazine before being passaged to naïve mice. Sulphadiazine basically rescues gametocytes by selectively killing the asexuals. In this experiment as there were no gametocytes, the population was entirely gametocyte non-producer type and hence succumb to the drug.

The parasites were passaged from each of the G424 and G426 lines into naïve mice (thus generating lines G424m1p1, G424m1p2, G426m1p1 and G426m1p2, respectively; “m” here denotes negatively selected and “p” denotes the number of times the line was passaged) before putting the lines G424, G425 and G426 on the drug Sulphadiazine (replacing 5-FC) on day 9 post-transfection. The passaged lines were put on 5-FC straightaway on the day of the passage. An equal amount of purified DNA (~100 ng) from each of the passaged lines was PCR-screened for successful integration of the complementation construct into the genome by amplifying various regions within and surrounding *PbAP2-G* (Table 6-5 and Figure 6-16). A summary of all the oligonucleotide sets used as primers in the PCRs, their details, expected length of bands, annealing temperature (Ta) settings used and extension time used for various PCR reactions are shown in Table 6-5.

Table 6-5

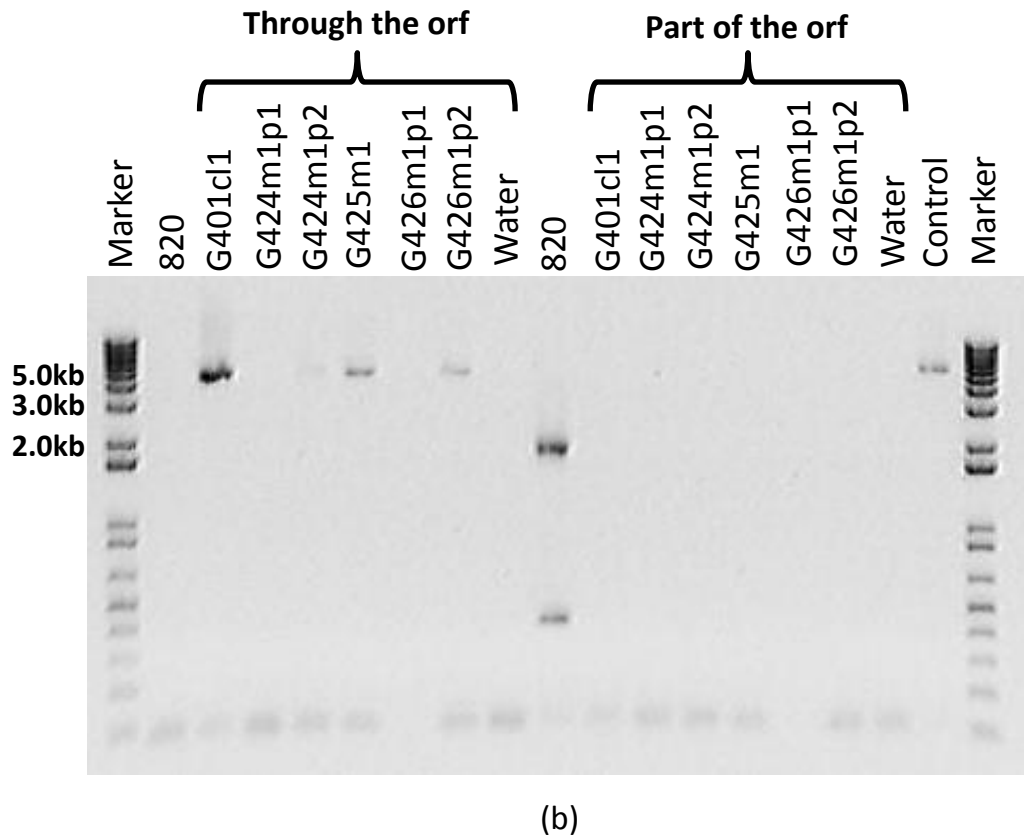
Oligo set	Details	Expected band size with DNA (kb)					Ta (°C)	Ext. Time (min)
		820	G401cl1	Transfectants	Water	HP		
1076/1132	5' integration check for ko	--	2.2	--	--	NA	54	5
53/1077	3' integration check for ko	--	2.6	--	--	NA	54	5
1076/1077	Through ORF	9	4.9	9	--	NA	54	8
1070/1239	Q1 ORF	1.9	--	1.9	--	NA	54	5
1233/1235	Control	NA	NA	NA	NA	5.9	54	8
1233/1242	Control	NA	NA	NA	NA	2.9	54	5

**Table 6-5. Basic PCR reaction settings for checking the 5' and 3' integration of the construct PbG01 into the G401cl1 genomic locus.** Apart from detecting integration, amplification of a part of the WT orf was also attempted to show that complementation was achieved.



(a)

**Figure 6-16 (a) PCR analyses of the DNA from the transgenic parasites resulting from the transfection of G401cl1 line with the complementation construct PbG01.** The transgenic lines (along with the WT line 820 as a control) were first tested for the absence of the knockout genotype by again performing the 5' and 3' integration checks. All the transgenic lines except G426m1p1 produced strong positive integration bands suggesting the persistence of the knockout genotype in the transfectants.



**Figure 6-16 (b) PCR analyses of the DNA from the transgenic parasites resulting from the transfection of G401cl1 line with the complementation construct PbG01.** All the lines (especially G426m1p1) were also screened for presence of the WT orf, fully or partially, by using appropriate oligos (Figure 6-16b). No WT orf signals could be obtained in any of the transgenic lines strongly suggesting the evidence was against successful complementation.

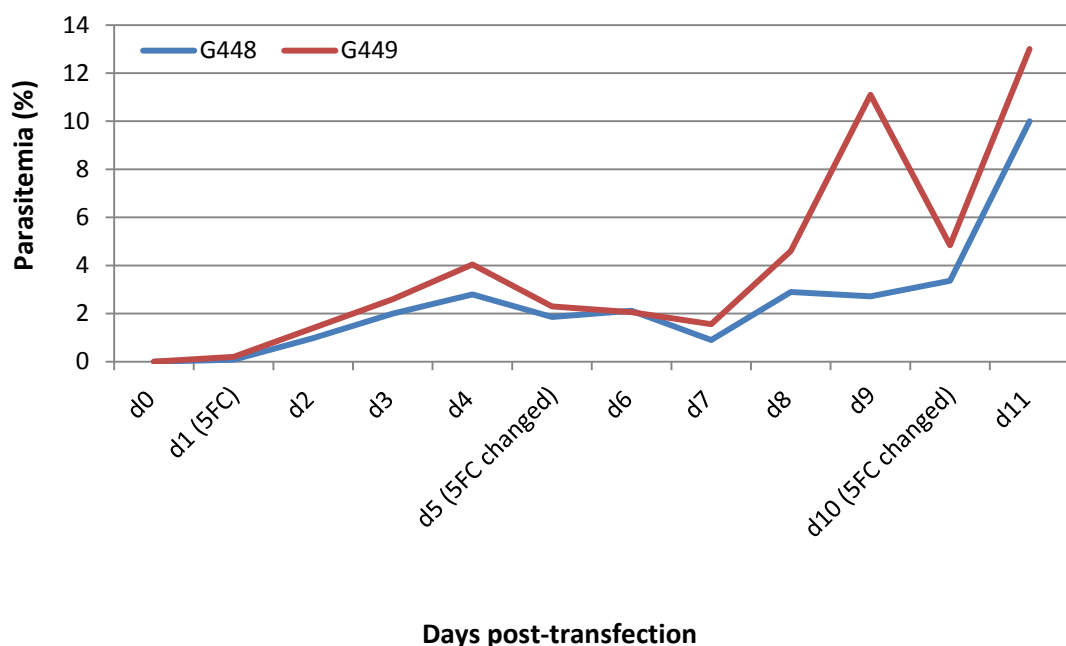
From the PCR analyses (Figure 6-16), there was no evidence of successful integration of the PbG01 construct into G401cl1 genome - the persistence of the knockout (5' and 3' integration positive) population in Figure 6-16(a) and absence of any evidence suggesting the amplification of a region of the *PbAP2-G* open reading frame from any of the transfectants indicating that complementation did not occur (Figure 6-16b). The genotypic findings perfectly matched with the continuous phenotypic examination of the lines G424m1p1-2, G425m1 and G426m1p1-2 by Giemsa-stained peripheral blood smear and flow cytometry. Both, the genotypic as well as phenotypic evidence were against any successful complementation of *PbAP2-G* in the transfected negatively selected lines. There were no "typical" gametocytes observed on either Giemsa-stained smears or FACS although some gametocyte-like parasites were noted on the smears very early during the course of negative selection.

Also, the negative selection kinetics (Figure 6-15) suggested that there was a parasite population that was NOT sensitive to 5-FC. This could happen in rather four possible ways - (a) when there are no parasites of the knockout genotype bearing the *yfcu* drug selectable marker which would imply a successful complementation attempt; or (b) a mutation has been acquired during the negative selection process which would confer resistance to 5-FC hence making the negative selection ineffective; the *yfcu* gene was sequenced from some of these transgenic lines and their sequence alignment with the WT *yfcu* gene did not reveal any sequence abnormality in *yfcu* (data not shown); or (c) the natural pro-GNP population selection pressure overwhelms the negative drug selection pressure. The latter two possibilities would naturally select for the GNP population (the knockouts) as they have a growth advantage over the gametocyte producer parasites and hence will ultimately overgrow them in time. The continuous rise in parasitemia during negative selection after an initial dip also suggests the latter possibility; (d) there is then the fourth possibility that the selection disadvantage offered by retaining a functional *yfcu* gene in the selection cassette could enforce the parasites to genetically rearrange in such a way that the selection cassette is removed from their genome, totally or partially. This was however, highly unlikely, as the PCR analyses in the majority of these transgenic lines were based on selection of at least one of the primers binding to the selection cassette, and *yfcu* in particular. Thus the presence of the expected sized band/s from any of these primers would imply the persistence of the unmutated selection cassette / *yfcu* in the transgenic parasite genomes. The presence of these bands almost every time such selection cassette based PCRs were performed ruled out this possibility of eliminating the cassette from the parasite genome.

Based on the remote possibility of having gametocytes in the transfectants (as suggested by Giemsa smears as well as the negative selection kinetics), the transfectants were put on sulphadiazine (on day 9 post-transfection; red vertical line in Figure 6-15) which selectively kills the asexual parasites leaving behind the gametocytes, if any. The transfectant lines were passaged and negative selection was maintained before being put on sulphadiazine. As seen in the latter half of the negative selection kinetics, the parasitemia sharply fell to zero after exposing the parasite population to

sulphadiazine. This implied that by the time the transfectants were put on sulphadiazine, the GNP population had already overgrown the gametocyte producers (if there were any to start with), thus showing that there was no complementation at all or if it was there, it could not be rescued in time.

Based upon the results of the first set of transfections to complement the full length *PbAP2-G* knockout line (G401cl1), a second set of repeat transfection experiments, in duplicate, was carried out with increased (up to double) the amount of PbG01 DNA used (5-10 µg per transfection) for transfection. The resultant transfectants, G448 and G449 were selected under 5-FC exactly as before (Figure 6-17, except the administration of sulphadiazine) and were examined by PCR analyses for integration (Table 6-6 and Figure 6-18), Giemsa-stained smears and flow cytometry for gametocyte production and hence complementation. The findings were similar to that mentioned in the first complementation experiment - no integration event on PCR analyses and no gametocytes on Giemsa smears and FACS.

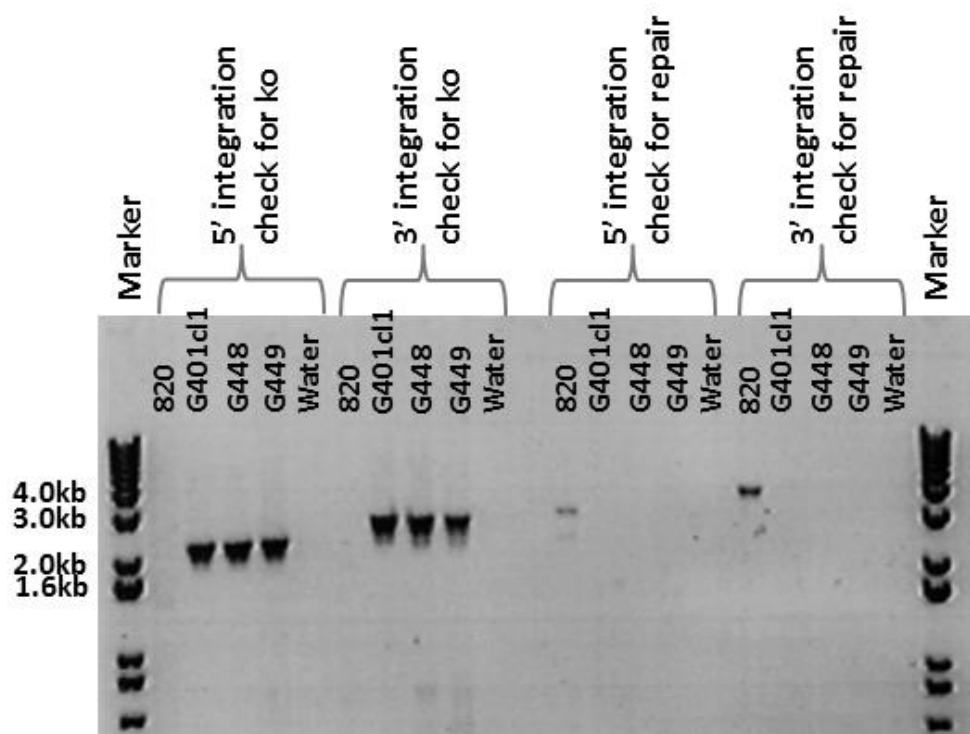


**Figure 6-17. Negative selection parasitemia kinetics for the transgenic lines G448 and G449 repaired (in the second round of transfections) by using the complementation construct PbG01 in the full length *PbAP2-G* knockout line G401cl1.** In contrast to the first round of transfections with the same construct (Figure 6-15), the drug 5-FC was started on the day of the transfection and was replaced with fresh drug at least every 4<sup>th</sup> day. Horizontal X-axis shows the progression of the experiment as days post-transfection and vertical axis shows the total parasitemia, as observed with the Giemsa-stained smears. Together with PCR analyses, Giemsa-stained smears, and flowcytometry, the kinetics in these lines suggest that the negative selection was not successful and thus the complementation attempt.

Table 6-6

Oligo set	Details	Expected band size with DNA (kb)					Ext. Time (min)
		820	G401cl1	G448-49	Water	HP	
1076/1132	5'int.check for ko	--	2.2	--	--	NA	4
53/1077	3'int.check for ko	--	2.6	--	--	NA	4
1057/1127	5'int.check for repair	2.7	--	2.7	--	NA	4
1128/1287	3'int. check for repair	3.4	--	3.4	--	NA	4
1070/1239	Q1 ORF	1.9	--	1.9	--	NA	4
1182/1240	Q2 ORF	1.8	--	1.9	--	NA	4
1241/1242	Q3 ORF	1.9	--	1.9	--	NA	4
1237/1071	Q4 ORF	2.1	--	2.1	--	NA	4
1231/1242	Control	NA	NA	NA	NA	3.5	4

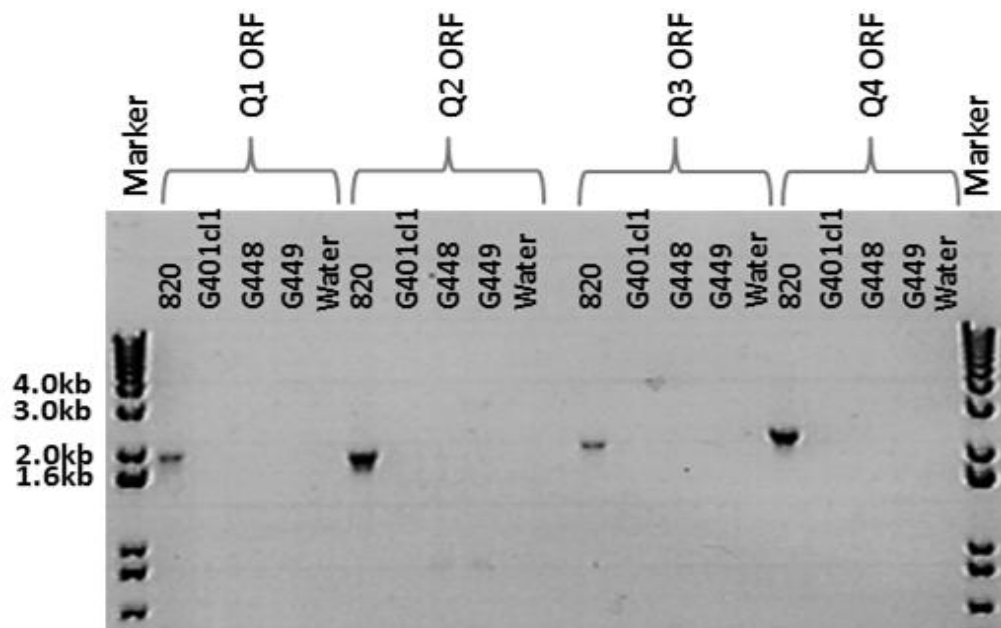
**Table 6-6. Basic PCR reaction settings for checking the 5' and 3' integration of the construct PbG01 into the G401cl1 genomic locus.** Apart from detecting integration, amplification of four parts of the whole WT *PbAP2-G* orf was also attempted to show that the complementation was achieved. The annealing temperature for the above reaction was set at 54°C.



(a)

**Figure 6-18 (a). PCR analyses of the DNA obtained from the transfectants following the PbG01 based complementation of the full AP2-G orf knockout line G401cl1.** The 5' and 3' integration check PCRs for the knockout lines and the repaired (complemented) lines are shown. Both the transfectants, G448 and G449 showed evidence of the 5' and 3' integration bands (2.2 kb and 2.6 kb, respectively) in favour of the presence of the knockout genotype (quite usual at this stage as these were not cloned lines) but the corresponding bands in favour of the presence of a repaired genotype (2.7 kb and 3.4 kb, respectively) were absent in both lines.

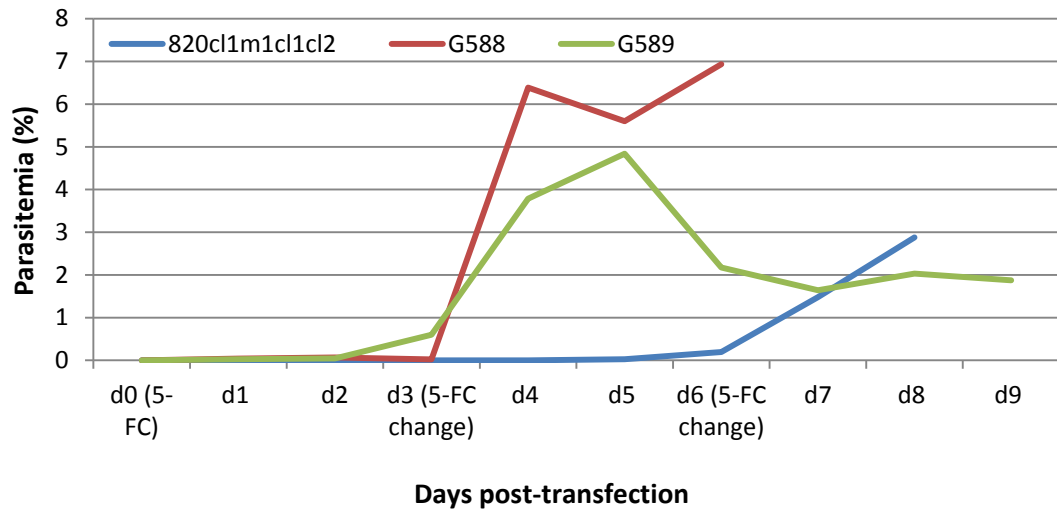




(b)

**Figure 6-18(b).** PCR analyses of the DNA obtained from the transfectants following the **PbG01 based complementation of the full *AP2-G* orf knockout line G401cl1**. No WT orf bands corresponding to the four regions of the *AP2-G* orf could be observed in the knockout line G401cl1 (expected) and in the lines G448 and G449 suggesting strongly that there has been no complementation of the *PbAP2-G* gene in the repaired full orf knockout variant.

Another set of transfection-negative selection experiments was done to retry complementation of *PbAP2-G* in G401cl1 using *PbG01* under enhanced negative selection pressure in view of the possibility of the natural GNP selection pressure overtaking the negative selection pressure. Thus, the transfectant lines G588 and G589 were generated in exactly the same way as before but changing the way the parasites were negatively selected. In addition to the usual dose of oral 5-FC, injectable 5-FC (1 mg per mouse per day, intraperitoneally) was supplemented to G588 and G589 infected mice continuously for 4 days to augment the negative selection pressure to counteract possible stronger pro-GNP natural selection pressure (Figure 6-19). Unfortunately again, no gametocyte producing population could be observed, neither in smears nor in flowcytometry, again suggesting no complementation had occurred.

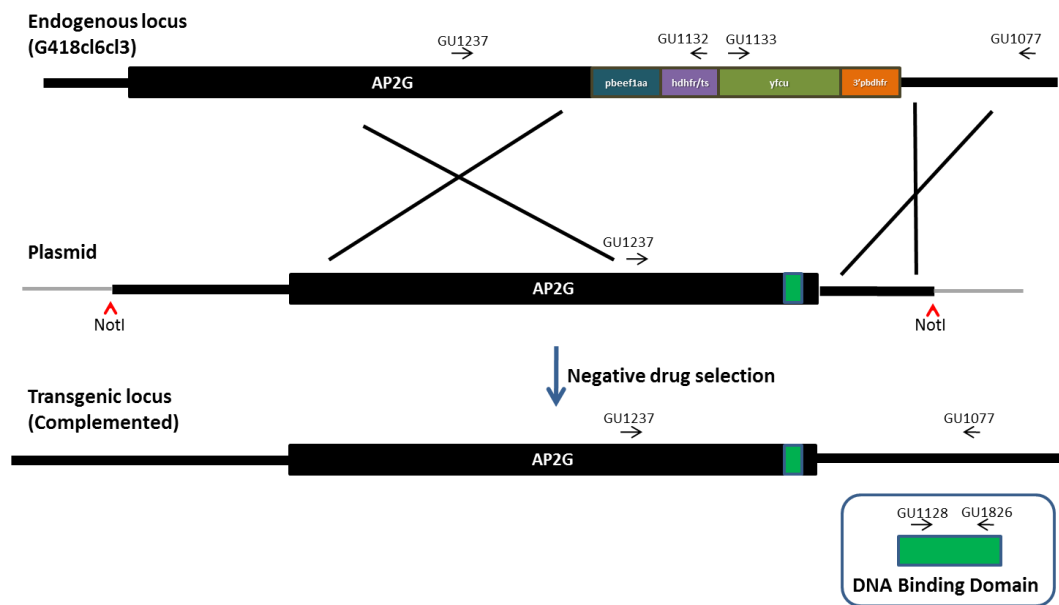


**Figure 6-19. Negative selection parasitemia kinetics for the transgenic lines G588 and G589 repaired (in the third round of transfections) by using the complementation construct PbG01 in the full length *PbAP2-G* knockout line G401cl1.** In contrast to the first two rounds of transfections with the same construct (Figures 6-15 and 6-17), the oral drug 5-FC was supplemented by injections of 5-FC (1 mg/mouse/day; intra-peritoneally) for four consecutive days starting from the day of transfection. The oral 5-FC was given as before. Horizontal X-axis shows the progression of the experiment as days post-transfection and vertical axis shows the total parasitemia, as observed with the Giemsa-stained smears. Together with PCR analyses, Giemsa-stained smears, and flowcytometry, the kinetics of these lines suggest that negative selection was not successful for the complementation attempt.

#### 6.2.2.1.2 Complementation in *PbAP2-G* DNA binding domain (DBD) knockout; G418cl6cl3

The complementation construct, *PbG01*, was utilized to repair the mutation/s present in the DBD knock out (from Chapter 5) as well (Figure 6-20). NotI digested (New England BioLabs) *PbG01* was transfected to G418cl6cl3 schizonts in quadruplet. The resulting transfectants (G547, G548, G590 and G591) were put on negative selection using the drug 5-FC 1.5 mg/ml in drinking water from day 0 post-transfection. For G590 and G591, in addition to the usual dose of oral 5-FC, injectable 5-FC (1 mg per mouse per day, intraperitoneally) was supplemented continuously for 4 days to augment the negative selection pressure to counteract stronger pro-GNP natural selection pressure. Oral 5-FC mediated negative selection pressure was tightly maintained (fresh 5-FC supplied every four days) and the parasitemia (including gametocytemia, if any) was monitored daily (as described in Materials and Methods) till day 14 post-transfection before culling the mice (Figure 6-22).

Diagnostic PCRs were performed on the DNA from negatively selected lines as per Table 6-7 and Figure 6-21. A summary of all the oligonucleotide sets used as primers in the PCRs, their details, expected length of bands, annealing temperature (Ta) settings used and extension time used for various PCR reactions are shown in Table 6-7.

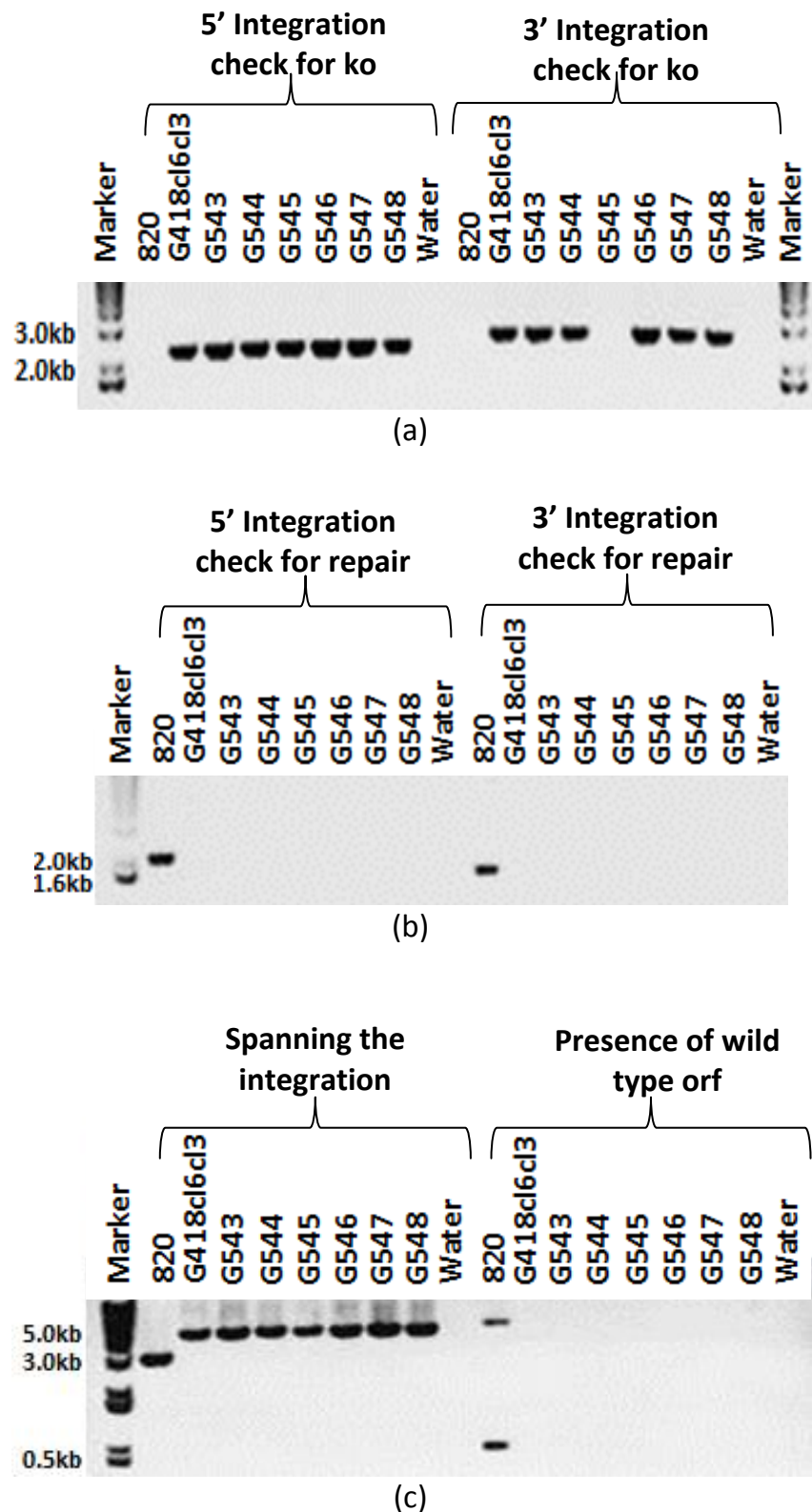


**Figure 6-20: Schematic (not-to-scale) showing the complementation strategy to repair the DBD knockout of *PbAP2-G* using full length WT copy of the gene as a recombineering based construct.** The endogenous locus for the DBD knockout (G418cl6cl3) had a selectable marker cassette (orange=3'UTR of *Pbdhfr*; dark blue=pbeef1aa; violet=hdhfr; green=yfcu) with only one copy of the 3' *Pbdhfr*/ts UTR making the cassette non-recyclable and hence usable for the negative selection-based repair (see text). The vector (PbG01) with two *NotI* linearization sites is shown. The likely areas for homologous recombination are shown as crosses. The transfectants were negatively selected using the drug 5-FC in drinking water. The resultant complemented transgenic locus is also depicted. Also shown are the relative position of various oligonucleotide primers used for confirmation of integration event (Table 6-7) and the resulting repair. The inset (bottom right corner) shows the magnified view of the DBD along with the PCR amplification oligonucleotide primer pair.

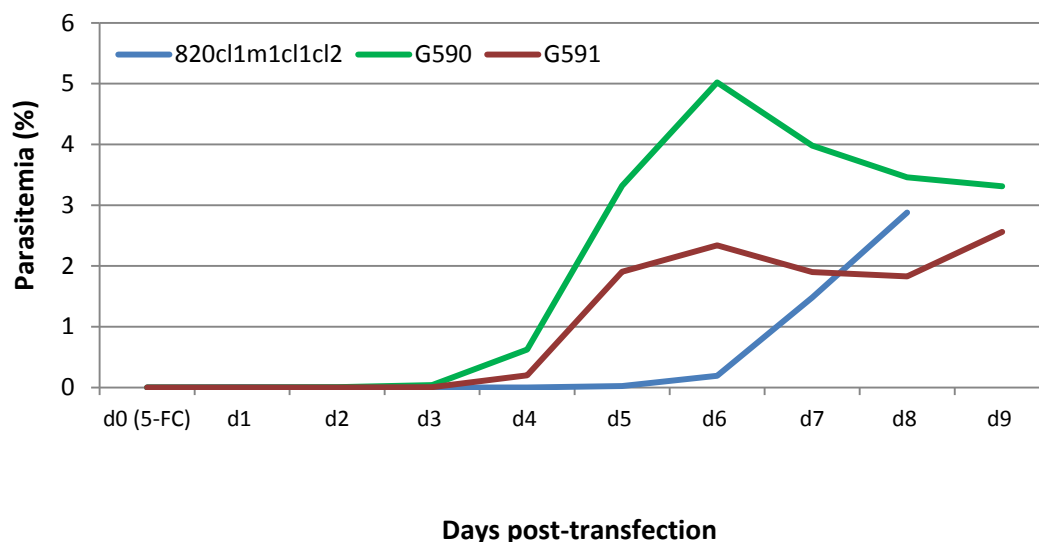
Table 6-7

Oligo set	Details	Expected DNA band size (kb)				Ext. Time (min)
		820	G418cl6cl3	G543-548	Water	
1237/1132	5'int. check for ko	--	2.2	--	--	4
1133/1077	3'int. check for ko	--	2.5	--	--	4
1237/1826	5'int. check for repair	2.0	--	2.0	--	4
1128/1077	3'int. check for repair	1.6	--	1.6	--	4
1237/1077	Spanning the integration	3.1	5.0	3.1	--	7
1128/1826	DBD ORF	0.5	--	0.5	--	4

**Table 6-7. Basic PCR reaction settings for checking the 5' and 3' integration of the construct PbG01 into the G418cl6cl3 genomic locus.** Apart from detecting integration, amplification of a part of the DBD WT *PbAP2-G* orf was also attempted to show that complementation was achieved. The annealing temperature for the above reaction was set at 54°C.



**Figure 6-21. PCR analyses of the DNA obtained from the transfectants following the *PbG01* based complementation of the *AP2-G* DBD ko line *G418cl6cl3*.** The 5' & 3' integration check PCRs for the knockout lines are shown in (a), both integration check PCRs for the repaired lines are shown in (b), and PCRs spanning the integration & amplifying a part of the DBD are shown in (c). All transfectants showed evidence of 5' and 3' integration (2.2 kb & 2.5 kb, respectively) except the 3' integration for ko in G545 denoting the presence of a ko genotype (usual as the lines were not yet cloned) but the corresponding bands denoting a repaired genotype (2.0 kb & 1.6 kb, respectively) were absent in all the transfectants. PCRs spanning the selectable marker cassette & amplifying the DBD orf (c) were also suggestive that the GNP population was still present in the transfectants. Further, no WT DBD orf bands could be observed in the repaired lines suggesting strongly that there was no complementation of the *PbAP2-G* DBD knockout variants.

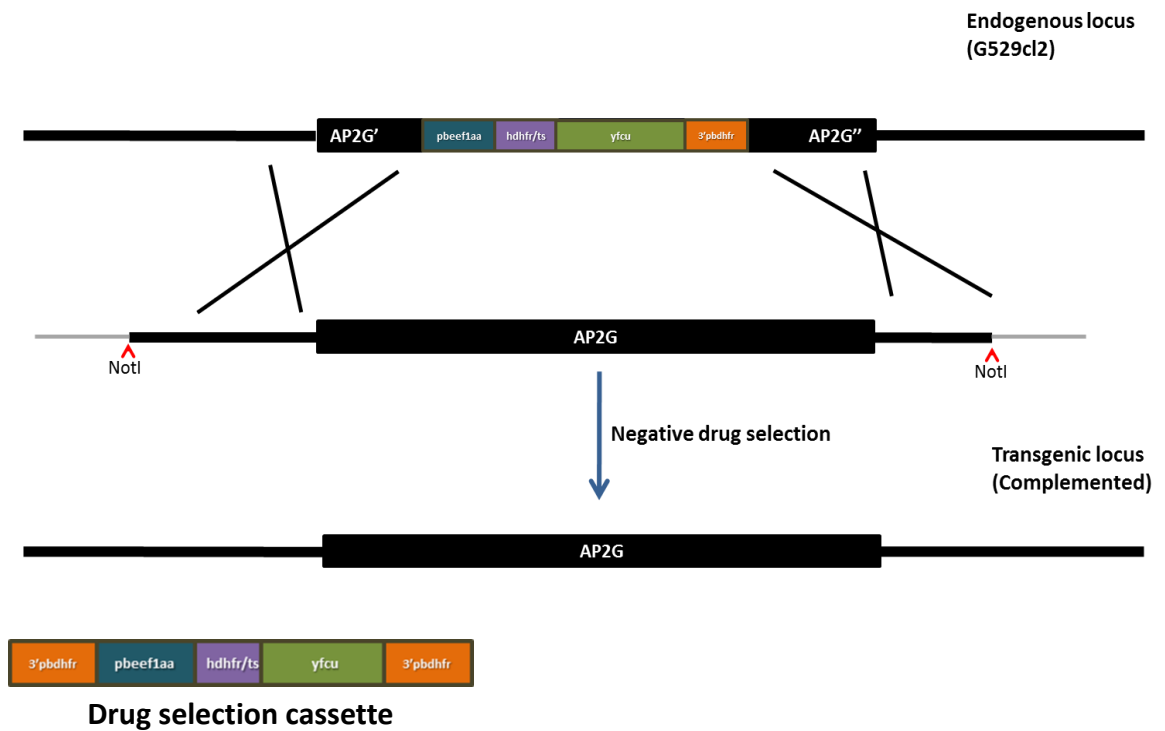


**Figure 6-22. Negative selection parasitemia kinetics for the transgenic lines G590 and G591 repaired (in the third round of transfections) by using the complementation construct PbG01 in the *PbAP2-G* DBD knockout line G418cl6cl3.** In contrast to the first two rounds of transfection with the same construct (Figures 6-15 and 6-17), the oral drug 5-FC was supplemented by injections of 5-FC (1 mg/mouse/day; intra-peritoneally) for four consecutive days starting from the day of transfection. The oral 5-FC was given as before. Horizontal X-axis shows the progression of the experiment as days post-transfection and vertical axis shows the total parasitemia, as observed with the Giemsa-stained smears. Together with PCR analyses (Figure 6-21), Giemsa-stained smears, and FACS, the kinetics in these lines suggest that the negative selection was not successful and so was the complementation attempt.

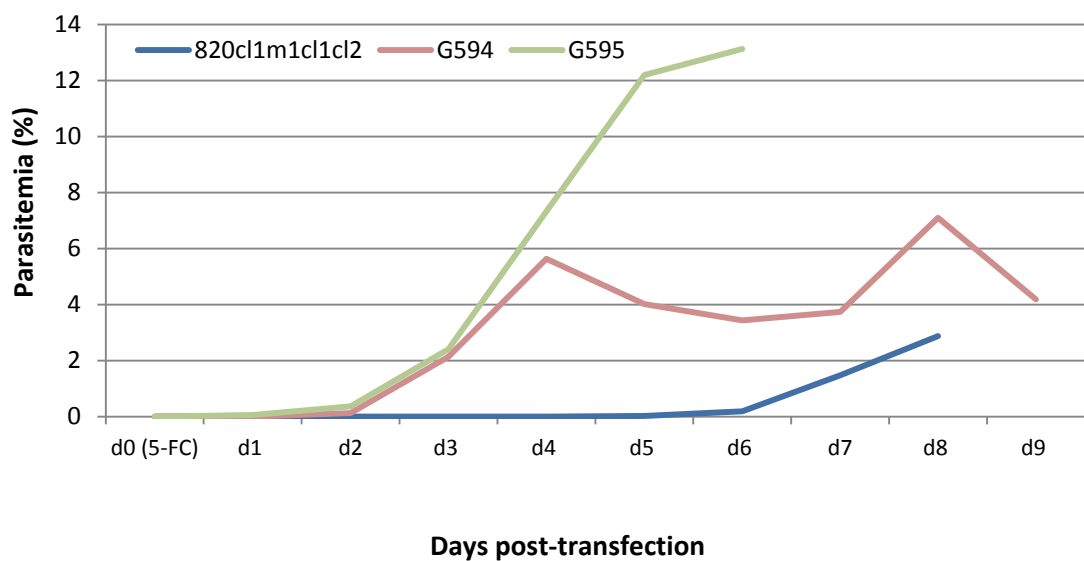
As evident from the PCR analyses (Figure 6-21a-c), there was no evidence of successful integration of the PbG01 construct into G418cl6cl3 genome - the persistence of the knockout (5' and 3' integration positive) genotype in Figure 6-21 (a) and absence of any evidence suggesting the amplification of a region of the DBD open reading frame from any of the transfectants indicating that the complementation did not occur (Figure 6-21b). The genotypic findings perfectly matched with the phenotypic examination of G590 and G591 during negative selection by Giemsa-stained peripheral blood smear and flow cytometry. Both, the genotypic as well as the phenotypic evidence were against any successful complementation of the DBD knockout in the transfected negatively selected lines.

### **6.2.2.1.3    *Complementation in PbAP2-G partial open reading frame knockout; G529cl2***

Genetic complementation of the *PbAP2-G* knockout variant line (G529cl2) which eliminated the part of the WT gene which had three naturally acquired GNP mutations was attempted with the PbG01 construct (Figure 6-23). NotI digested (New England BioLabs) *PbG01* was transfected to G529cl2 schizonts in duplicate. The resulting transfectants (G594 and G595) were put on negative selection using the drug 5-FC 1.5 mg/ml in drinking water from day 0 post-transfection. In addition to the usual dose of oral 5-FC, injectable 5-FC (1 mg per mouse per day, intraperitoneally) was supplemented continuously for 4 days to augment the negative selection pressure to counteract any stronger pro-GNP natural selection pressure. Oral 5-FC mediated negative selection pressure was tightly maintained (fresh 5-FC supplied every four days) and the parasitemia (including gametocytemia, if any) was monitored daily (please see section 2.1.5 in Materials and Methods) till day 9 post-transfection before culling the mice (Figure 6-24). The parasitemia kinetics under negative selection (the spurt in proliferation at day 2 post-transfection; Figure 6-24) suggests that the negative selection was not efficient in removing parasites bearing the *yfcu* gene. Together with absence of visible gametocytes in Giemsa-stained peripheral smears and FACS, these findings suggested a failure of complementation of the *PbAP2-G* partial orf knockout and hence PCR analyses for the integration event were not performed.



**Figure 6-23. Schematic (not-to-scale) showing the complementation strategy to repair the partial knockout of *PbAP2-G* which eliminated the region of the orf (which had the three naturally acquired GNP mutations) using full length WT copy of the gene as a recombineering based construct (PbG01).** The endogenous locus for the partial *AP2-G* knockout (G529cl2) had a selectable marker cassette (orange=3'UTR of *Pbdhfr*; dark blue=pbeef1aa; violet=hdhfr; green=yfcu) with only one copy of the 3' *Pbdhfr*/ts UTR making the cassette non-recyclable and hence usable for the negative selection-based repair (see text). The vector (PbG01) with two *NotI* linearization sites is also shown. The likely areas for homologous recombination are shown as crosses. The transfectants were negatively selected using the drug 5-FC in drinking water. The resultant complemented transgenic locus is also depicted.



**Figure 6-24. Negative selection parasitemia kinetics for transgenic lines G594 and G595 (in the third round of transfections) by using the complementation construct PbG01 in the *PbAP2-G* partial orf knockout line G529cl2.** In contrast to the first two rounds of transfections with the same construct (Figures 6-15 and 6-17), the oral drug 5-FC was supplemented by injections of 5-FC (1 mg/mouse/day; intra-peritoneally) for four consecutive days starting from the

day of transfection. The oral 5-FC was given as before. The X-axis shows the progression of the experiment as days post-transfection and the axis shows the total parasitemia, as observed with Giemsa-stained smears. Together with absence of gametocytes in Giemsa-stained smears and FACS, the kinetics in these lines suggests that the negative selection was not successful and so nor was the complementation attempt.

### 6.2.2.2 Complementation of the knockouts using a mutation-free PCR product

Complementation using a simple PCR product is based upon the “GIMO approach” developed recently to recycle the drug selectable marker by homologous recombination (Lin et al., 2012). The “Gene In/ Marker Out” or the “GIMO” approach simply requires a PCR-amplified wild type copy of the gene or the part of the gene to be restored, including upstream and downstream arms homologous to the target genome and the presence of a negative selection marker contained in the mutant to be complemented. Since the *PbAP2-G* knockouts had the negative selection marker in the genome, the GIMO-based complementation approach was used. The full-length knockout variant of *PbAP2-G* (G401cl1) was not suitable for this technique as it was difficult to obtain a single PCR product spanning the whole length of the gene (7.02 kb) plus a kilobase of homology arms on each side, thus making the of a PCR product of ~9 kb in sufficient amount was considered practically impossible.

Thus the GIMO-based complementation approach was attempted only with the DBD knockout line (G418cl6cl3; Figure 6-25). The PCR product for the repair was amplified as two variants (short and long) from the WT line 820, using primer sets GU1389/GU1514 for the shorter product (~2 kb) and GU1416/GU1059 for the longer product (~2.8 kb) using the HiFi Taq-polymerase (Figure 6-25).

Approximately 5 µg each of the purified PCR products (long and short) were transfected to G418cl6cl3 schizonts - the short one was transfected twice whereas the long one was transfected four times. The resulting transfectant lines (G543 and G544 for short and G545, G546, G592 and G593 for long product) were put on negative selection using the drug 5-FC (1.5 mg/ml) in drinking water from day 0 post-transfection. For G592 and G593, in addition to the usual dose of oral 5-FC, injectable 5-FC (1 mg per mouse per day, intra-peritoneally) was supplemented continuously for 4 days to augment the negative selection

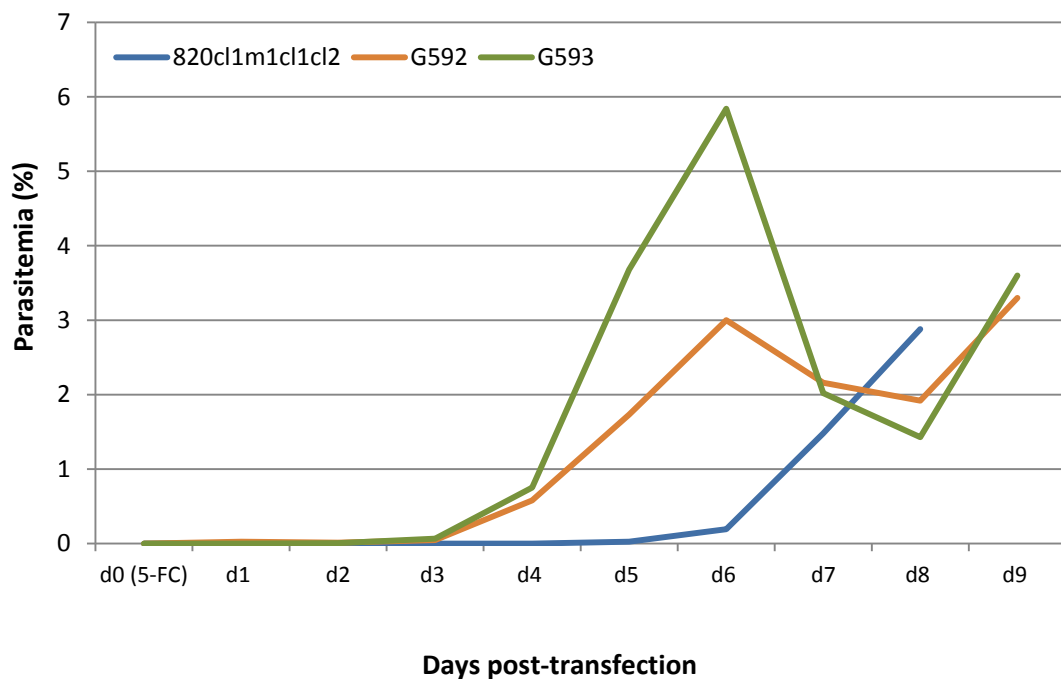


pressure to counteract any stronger pro-GNP natural selection pressure. Oral 5-FC mediated negative selection pressure was tightly maintained (fresh 5-FC supplied every four days) and the parasitemia (including gametocytemia, if any) was monitored daily (please see section 2.1.5 in Materials and Methods) till day 14 post-transfection before culling the mice (Figure 6-26).



**Figure 6-25. Schematic (not-to-scale) showing the PCR-based complementation strategy to repair the DBD knockout of *PbAP2-G* using long and short PCR products amplified from WT *PbAP2-G* and bearing upstream and downstream recombinable homology arms.** The WT *PbAP2-G* endogenous locus is shown at the top with the DBD (green vertical bar) and the region around the domain which was knocked out (red transparent box). The oligonucleotide pairs used to amplify the long and short PCR products, GU1416/GU1059 and GU1389/GU1514, respectively are also indicated. The endogenous locus for the DBD knockout (G418cl6cl3) had a selectable marker cassette (orange=3'UTR of Pbdhfr; dark blue=pbeef1aa; violet=hdhfr; green=yfcu) with only one copy of the 3' Pbdhfr UTR making the cassette non-recyclable and hence usable for negative selection-based repair (see text). The long (2.8 kb) and short (2.0 kb) PCR products are also shown above the DBD knockout locus. The likely areas for homologous recombination are shown as crosses. The transfectants were negatively selected using the drug 5-FC in drinking water. The resultant complemented transgenic locus is also depicted. Also shown are the relative position of various oligonucleotide primers used for confirmation of an integration event (Table 6-7) and the resulting repair. The inset (bottom right corner) shows a magnified view of the DBD along with its PCR amplification oligonucleotide primer pair.

Diagnostic PCRs were performed on purified DNA from the negatively selected lines G543, G544, G545, and G546 as per Table 6-7 and Figure 6-25 and shown in Figure 6-21. A summary of all the oligonucleotide sets used as primers in the PCRs, their details, expected length of bands, annealing temperature (Ta) settings used and extension time used for various PCR reactions are shown in Table 6-7.



**Figure 6-26. Negative selection parasitemia kinetics for the transgenic lines G592 and G593 repaired (in the third round of transfections) by using the PCR based complementation products in the *PbAP2-G* DBD knockout line G418cl6cl3.** In contrast to the usual negative selection of transfectants, the oral drug 5-FC was supplemented by injections of 5-FC (1 mg/mouse/day; intra-peritoneally) for four consecutive days starting from the day of transfection. The oral 5-FC was given as before. Horizontal X-axis shows the progression of the experiment as days post-transfection and vertical axis shows the total parasitemia, as observed with the Giemsa-stained smears. Together with absence of gametocytes in Giemsa-stained smears and FACS, the kinetics in these lines suggests that the negative selection was not successful and so neither was the complementation attempt.

The parasitemia kinetics of lines G52 and G593 under negative selection (the spurt in proliferation at day2 post-transfection; Figure 6-26) suggests that the negative selection was not efficient in removing parasites bearing the *yfcu* gene. Together with absence of visible gametocytes in Giemsa-stained peripheral smears and FACS, these findings suggested a failure of complementation of the *PbAP2-G* partial orf knockout and hence PCR analyses for the integration event were not performed for these lines.

## 6.3 Summary and discussion

Genetic complementation, as such, is the gold standard experiment to establish and prove a causal hypothesis generated during forward and reverse genetic manipulations. Complementation of a mutated gene in *Plasmodium* presents specific problems both in terms of the relative paucity of the number of drug selectable markers available to perform successive genetic modification at a particular genetic locus and also in terms of the extreme AT-richness of the *Plasmodium* genome.

Forward and reverse genetic studies, Chapter 3 and Chapter 5, respectively, identified a single gene *PbAP2-G* (PBANKA\_143750) as the master regulator of the process of commitment to gametocytogenesis in *P. berghei*. *PbAP2-G* natural and null mutants did not produce any gametocytes and therefore the resultant GNP population had a significant growth advantage over the gametocyte producer parasites, sufficient enough to overgrow the latter in a mixed population. In this particular study, complementation efforts were further challenged by at least two more factors - first, the length of the gene to be complemented was ~7 kb thus giving rise to inherent problems associated with cloning and manipulation of such a large, and single DNA fragment; and second, the reversal of the GNP phenotype to a gametocyte-producer phenotype had to happen against a strong natural selection pressure in favour of the GNP parasites, posing problems in negatively selecting the transgenic complemented parasites.

Despite the above mentioned deterrents to successful complementation of *PbAP2-G* null mutants, there were two distinct advantages: the natural GNP mutants did not have the drug selectable marker cassette as they were selected using a forward genetics screen and hence the need for an additional selectable marker or for marker recycling was not required. But at the same time, this generated the need of inserting a positive selection marker into the complementation construct to positively select for the transgenic complemented parasites. This drug selectable marker cassette had to be incorporated into the WT genome during complementation which created specific problems (discussed below). The second advantage was the availability of recombineering-based

vectors with relatively longer homology arms for more efficient homologous recombination events.

Two types of genetic mutants were generated during the study which needed complementation data to prove the role of PbAP2-G in commitment to gametocytogenesis. The first type of mutants were generated by serial passage of ten isogenic wild type gametocyte producer lines (WT 820) thereby producing three natural GNP mutants, m9, m7 and m8. The PbG01\_COMP recombineering-based complementation construct which was used to repair the mutations was available in two varieties - one with the drug selectable marker cassette was incorporated 1.2 kb upstream to the *AP2-G* translational start site (PbG01\_COMP-UP) and the other with the cassette sitting ~2 kb downstream to the *AP2-G* stop codon (PbG01\_COMP-DOWN). The functional restoration of the GNP mutants to WT was hampered by the possible interruption of the *AP2-G* promoter while using the PbG01\_COMP-UP construct. This might have lengthened the promoter disturbing the nucleosomal phasing and hence a possible hindrance in the epigenetic regulation of the *AP2-G* expression.

Evidence of successful complementation (although incomplete) was seen using the other construct, the PbG01\_COMP-DOWN. But the number of gametocyte producer parasites generated through successful functional restoration of *AP2-G* was quickly overgrown by the persistent remaining GNP parasites in the transgenic complemented line. With time and even minimal passage, these transgenic lines rapidly generated the GNP phenotype by out competing the gametocyte producer parasites even before phenotype rescue experiments like cloning and mosquito passage were carried out. However, the experiments were repeated and refined with more precise and timely rescue of the gametocyte producer population, and functional complementation of all the natural GNP mutants was finally achieved (work carried out by Rachael Orr after I finished my lab work, Andy Waters' group, personal communication; data not shown).

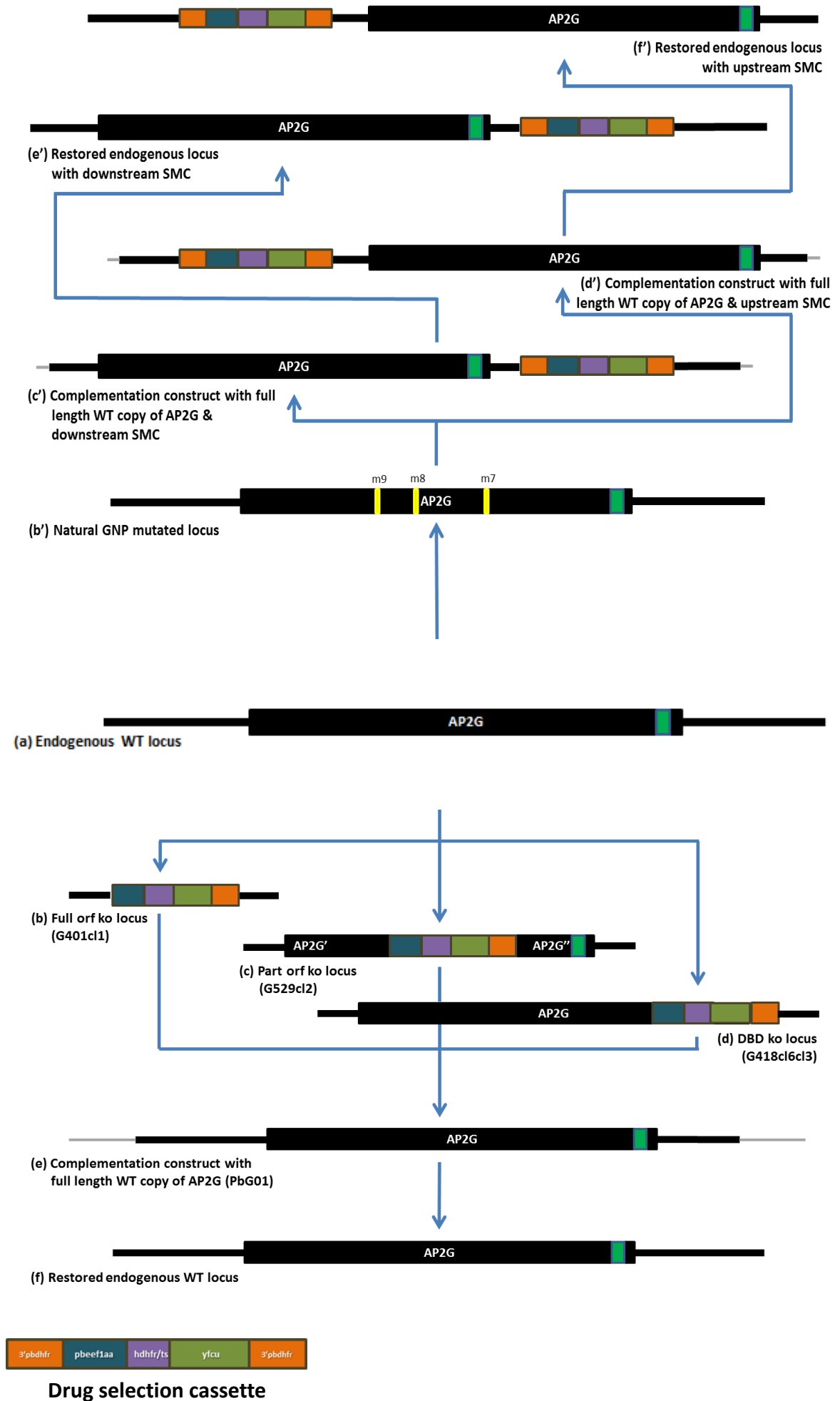
The second type of GNP mutant was engineered using reverse genetics techniques (gene knockout using homologous recombination). Three length variants of *PbAP2-G* knockouts were created - full orf knockout (G401cl1),

partial knockout of the region of the *AP2-G* orf that bears the 3 mutations identified for the naturally selected GNP lines (G529cl2) and a partial orf knockout eliminating the *PbAP2-G* DBD (G418cl6cl3). Complementation of these knockout variants was attempted using two approaches both using negative selection - full length wild type copy of the gene (recombineering-based construct, PbG01) which was applicable to all three knockout variants and partial length wild type copy of the gene (PCR-based construct) which was applicable to the latter two knockout variants. None of the approaches could restore the phenotype in any of the engineered mutants despite multiple experiments using a variety of modifications. At least two possible reasons could be involved either singly or in combination which resulted in the failure to complement - ineffective negative selection which could have been due to an undetected mutation in the *yfcu* gene conferring resistance to 5-FC or site-specific genetic rearrangements making the selectable marker cassette non-functional, and a stronger natural phenotypic selection favouring the GNP population. Figure 6-27 and Table 6-8 summarise the multitude of the techniques used to complement the gene in the available mutants and the results thereof.

Table 6-8

<i>PbAP2-G</i> mutant type	Line	Complementation constructs used			
		PbG01	PCR-product	PbG01_COMP -UP	PbG01_COMP-DOWN
Complete orf ko	G401cl1	Failure	Not attempted	Not Applicable	
Partial orf ko	G529cl2	Failure	Failure		
DBD ko	G418cl6cl3	Failure	Failure		
Natural GNP	M9, M8 & M7	Not Applicable		Failure	<b>SUCCESS</b>

**Table 6-8. A comprehensive summary table showing the outcomes of different types of complementation constructs used in the study to functionally restore a range of *PbAP2-G* null mutants thus diversifying the choice and applicability of methods.** As reasoned out in the main text, the GIMO/PCR-based repair was not attempted for complementing the complete orf knockout line G401cl1 because of the obvious reasons hampering the amplification of such a long PCR product. Neither the recombineering-based construct bearing the full length functional copy of WT *PbAP2-G* (PbG01) nor the PCR-based long/short products could complement the gene in any of the three length variants of the *AP2-G* knockouts. In the naturally acquired GNP lines, the mutations were successfully repaired using the recombineering based construct bearing the full length functional copy of WT *PbAP2-G* with the drug selectable marker cassette placed downstream to the *AP2-G* stop codon (PbG01\_COMP-DOWN).



**Figure 6-27. A comprehensive schematic cartoon (not-to-scale) to summarise the range of experiments used in the study to complement the *PbAP2-G* in the three naturally acquired GNP mutants (top panel) and the three length knockout variants (bottom).** Note: The PCR-based complementation approach to repair the knockouts is not shown here to preserve the simplicity and clarity of the diagram. The WT *PbAP2-G* endogenous locus from the line 820 is shown in the centre (a). Above the WT locus are depicted the various complementation approaches used to repair the three natural GNP mutants (b' through f') and below the endogenous locus are shown the same for the three knockout variants (b through f). The full length positive-negative drug selectable marker cassette (SMC) is shown at the bottom right of the figure as multi coloured rectangular bar (orange=3'UTR of *Pbdhfr*; dark blue=*pbeef1aa*; violet=*hdhfr*; green=*yfcu*). Black thick bar labelled *AP2-G* = open reading frame of *PbAP2-G*; *AP2-G'* and *AP2-G''* on *AP2-G* orf = the upstream and downstream remainder of the *AP2-G* wherein the middle portion of the gene is eliminated and replaced by the SMC (Figure 6-27c); Green vertical rectangle on *AP2-G* orf = DNA binding domain; multi-coloured rectangle = SMC; yellow vertical bars on *AP2-G* = three naturally acquired *PbAP2-G* mutations

Other possible approaches which could have used to complement the defect include a single crossover (SCO) homologous recombination based repair of the natural GNP mutants which would restore at least one functional copy of the wild type *AP2-G*, expression of the gene under its own promoter from an exogenous locus such as the 230p locus and a Cre-loxP based generation of knockout followed by recycling of the selectable marker cassette and a subsequent positive selection based introduction of the wild type copy of the gene into the endogenous locus. The SCO based approach was attempted along with the GFP/HA tagging of *PbAP2-G* thus generating a repair-tagged vector for complementation and tagging. The vector was not processed and transfected due to paucity of time and simultaneous generation of an independent tagged line and successful complementation of the natural GNP mutants using the *PbG01\_COMP-DOWN* construct (see above). Expressing *PbAP2-G* from an exogenous locus is an attractive option but then the length of the gene along with its endogenous promoter and possibly 3'UTR to be cloned remains a possible deterrent to using such an approach.

It would also be interesting to study the effects of complementing an orthologous copy of the *PbAP2-G* from *P. falciparum* (or even trans-genera complementation, such as, from *Theileria spp.*; Huynh et al., 2004; Zou et al, 2009) and analyse the functional restoration of the gene. Keeping in mind the high degree of conservation of the functional DNA-binding domains as well as their recognition motif within *apicomplexa*, a successful functional complementation in such a case would at least establish the proof of principle - the conservation of the domain and motif is functional. Because the regions out

with the DBD are not thought to be conserved between orthologous *AP2-G*, it would also be intriguing to see if there is any role of these regions in controlling the AP2-G downstream actions. That would explain how the mutation in GNP line 233 could generate the GNP phenotype and provide a possible role of protein-protein or other interactions by this or other unidentified domains encoded by the *PbAP2-G* gene.



## **7      Analysing protein-DNA interactions**

## 7.1 Introduction

Identification and subsequent binding of a class of proteins called transcription factors (TF) to their cognate *cis*-elements in promoters and other regulatory regions in the genome comprise the initial steps in controlling the process of transcription of specific genes in eukaryotes (Phillips & Hoopes, 2008). TFs work in a close orchestration with epigenetic control (alteration of the chromatin structure) of the DNA containing the recognition motif for these TFs and thus regulate the complexity of gene expression in a combinatorial manner (Lemon and Tjian, 2000; Zaret and Carroll, 2011). Transcription factors contain a DNA-binding domain (DBD) which recognizes a specific recognition motif (usually 6-10 base pairs long) in a sequence-specific manner (Phillips & Hoopes, 2008). Any changes in TFs activity and/or regulatory specificity, for example, a deletion of the TF encoding gene or mutations in the TF binding sites (TFBS), is highly likely to influence phenotypic diversity and ecological adaptation among the organisms affected (Bustamante et al., 2005; Wray, 2007; Lopez-Bigas et al., 2008; De et al., 2008). A comprehensive understanding of the TFs, the TFBS and their interactions is thus crucial in attributing TFs for disease etio-pathogenesis and phenotypic variability.

Documenting or establishing the interaction between the DBD and its target DNA sequence is crucial for dissecting the downstream role of the TFs in terms of the specificity of action and the potential target genes for the TF. *In vitro* techniques such as Electrophoretic Mobility Shift Assay or EMSA are an ideal way to identify whether TFs (or their DBDs) interact with the TFBS or not (Elnitski et al., Genome Res 16; 2006). It has already been established that the *Pf AP2-G* DBD recognizes and subsequently binds to a specific 4-base pair core target sequence nnGTACnn *in vitro*. This target sequence is bioinformatically predicted to be highly conserved across orthologs of the *AP2-G* throughout the *apicomplexa* (Campbell et al., 2010). This predicted conservation of the core target sequence provides a basis of examining the interaction of the *AP2-G* DBD with the conserved DNA sequence *in vitro*. The aim was to identify the presence of a sequence-specific interaction of the DBD and the identified target sequence thus enabling us to predict possible downstream target genes of *PbAP2-G* based on the presence of the target motif in their 5'upstream region.

## 7.2 GST tagged DNA Binding Domain (DBD)

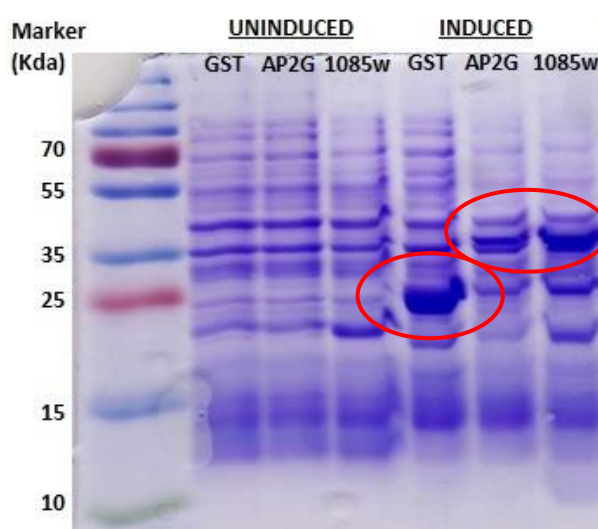
To study the binding interaction between the DNA Binding Domain (DBD) of PbAP2-G with the cognate DNA sequence, N-terminal Glutathione S-transferase (GST) tagged AP2 DBDs were used. GST on its own and N-terminal GST fused to extended ApiAP2 DNA binding domains (all cloned into pGEX-4T1) from *P. falciparum* AP2-G (PFL1085w) and *P. berghei* AP2-G (PBANKA\_143750) were kindly gifted by Manuel Llinas, Bjorn Kafsack and April Williams (Princeton University, USA). Extended versions of the AP2-G DBDs were used for molecular cloning which contained the DBD along with some of flanking residues that were likely to be important for domain stabilization (Campbell et al., 2010). GST protein was used both as expression control as well as the negative control for assessing binding of the AP2-Gs and the cognate DNA sequence. *Pf* AP2-G served as a positive control for EMSA as it has already been shown to bind to a conserved DNA sequence in a sequence-specific manner (Campbell et al., 2010; Manuel Llinas, personal communication). The target proteins were expressed in Rosetta (DE3) pLys S competent cells with 0.2 mM IPTG at 25°C and batch-purified using affinity chromatography (Glutathione HiCap Matrix slurry; Qiagen). The purity of protein was estimated by 10% SDS PAGE and the eluted proteins were quantified by spectrophotometry, absorbance at 260 nm. The eluted protein was concentrated and buffer exchanged using an Amicon Ultra-0.5 Centrifugal filter; 30K device (Milipore).

The expected size (in KDa) of GST and the two N-terminal GST-tagged DBDs is shown in Table 7-1. The successful induction of protein expression using 0.2 mM fresh IPTG is shown in Figure 7-1, whereas the migration of the purified tagged proteins in 10% SDS-PAGE is shown in Figure 7-2. As evident from Figure 7-2 (a & b) the amount of purified GST tagged proteins was adequate to be detected by SDS-PAGE for GST and *Pb* AP2-G. Although the amount of purified *Pf* AP2-G GST::DBD was minimal (as evident by faint bands in Figure 7-2(c)), it was sufficient to show the interaction with the cognate motif in EMSAs (see below).

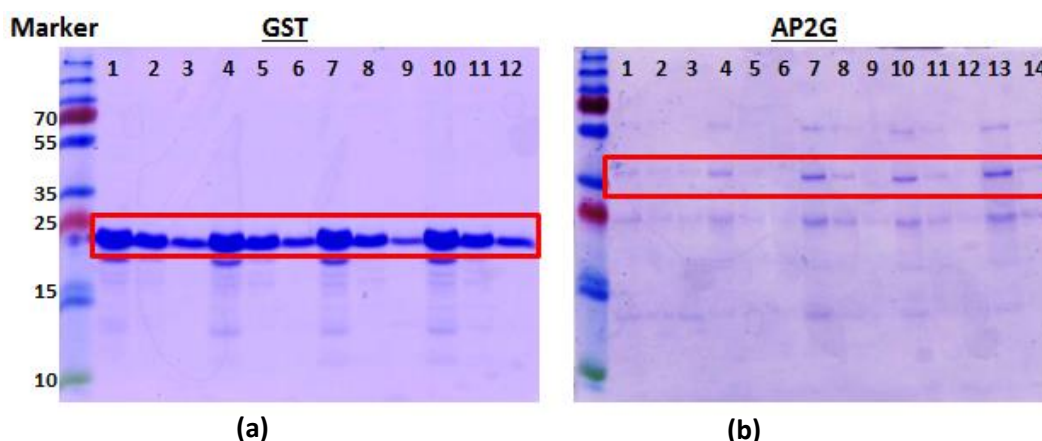
Table 7-1

Protein	Gene id	DBD_ext MW (KDa)	Tag MW (KDa)	DBD::tag MW (KDa)
<i>Pf</i> AP2-G DNA binding Domain	PFL1085w (PF3D7_1222600)	14.7	26 (GST)	40.7
<i>Pb</i> AP2-G DNA binding Domain	PBANKA_143750	14.3	26 (GST)	40.3

**Table 7-1. Relative expected sizes of N-terminal GST tagged AP2-G DNA binding domains and GST on its own.** The size of the AP2-G DBD shown here are that of the extended versions i.e. DBD with few flanking residues. Both the extended AP2-G::GST proteins are ~40 KDa in size.

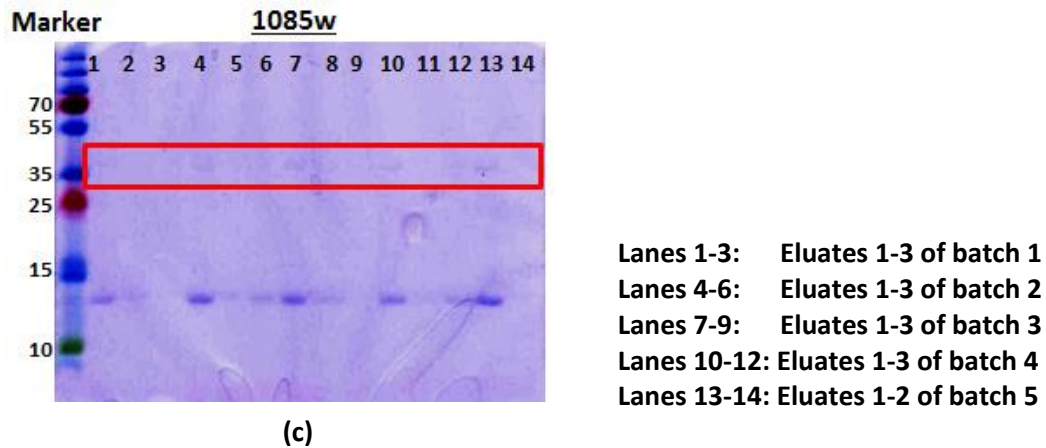


**Figure 7-1. A 10% SDS-PAGE image showing the separation of bands from the uninduced lysates and lysates after inducing the cultures with 0.2 mM fresh IPTG.** The gel was stained with Coomassie Blue for ~40 minutes and destained overnight with freshly prepared destaining solution. GST = Glutathione-S-Transferase; AP2-G = N-terminal GST tagged *Pb* AP2-G (PBANKA\_143750) DBD; and 1085 = N-terminal GST tagged *Pf* AP2-G (PF11\_1085w) DBD. Bands of desired size are shown in red circles.



(a)

(b)



**Figure 7-2. A 10% SDS-PAGE image showing the separation of bands from the batch-purified eluates** – three sequential eluates from 5 batches of protein purification (except in batch 5, where only 2 sequential eluates were available). The gel was stained with Coomassie Blue for ~40 minutes and destained overnight with a freshly prepared destaining solution. (a) GST = Glutathione-S-Transferase; (b) AP2-G = N-terminal GST tagged *Pb* AP2-G (PBANKA\_143750) DBD; and (c) 1085w = N-terminal GST tagged *Pf* AP2-G (PF11\_1085w) DBD. Bands of the desired size are shown in red rectangles. Purified proteins of the desired size were obtained from almost all eluates of all batches for GST and *Pb* AP2-G GST::DBD. Very small amount of *Pf* GST::AP2-G was obtained as evident in Figure 7-2(c). However, the amount was sufficient to perform EMSAs (see below).

## 7.3 AP2-G DBD binds to the expected DNA sequence

### 7.3.1 *The DBD recognizes a DNA sequence / motif: experimental setup*

It has already been shown by Protein Binding Microarray (PBM) experiments in *Plasmodium falciparum* that each of the different ApiAP2 transcription factors recognize and bind to at least one specific DNA sequence (motif) through their DNA binding domain/s (DBDs; Campbell et al., 2010). For PF11\_1085w, the *Pf* ortholog of PBANKA\_143750 (AP2-G), this target motif consists of GTAC as a core region and flanked by G/A, T and A, C/T on either side (G/A)TGTACA(C/T). Since the AP2 domains are highly conserved throughout the *Plasmodium* spp., the DBDs of the orthologous AP2 TFs were also predicted to bind to the same sequence across all *Plasmodium* spp. (Campbell et al., 2010).

Because of the DBDs of all known ApiAP2's are predicted to be highly conserved across *apicomplexa*, the implication that the *Pb* AP2-G DBD should bind to the same motif as recognized by the *Pf* ortholog was utilized to assess *in vitro* binding of the GST tagged *Pb* AP2 DBD to the motif sequence (G/A)TGTACA(C/T) by EMSA. The target recognition sequence was synthesised in two ways - as commercially synthesised 62-bp hypothetical oligonucleotides containing the core conserved recognition motif nnGTACnn flanked by an upstream (AATATTATAATAGTCGTAGCCATCAAT) sequence and a downstream (ATGGTAATATAGATTTTCGTTTATATT) sequence. The core sequence was modified/mutated retaining the same flanking sequences to show that the binding is sequence specific and that any single-base change in the core sequence abolishes the binding of the DBD to its cognate DNA sequence.

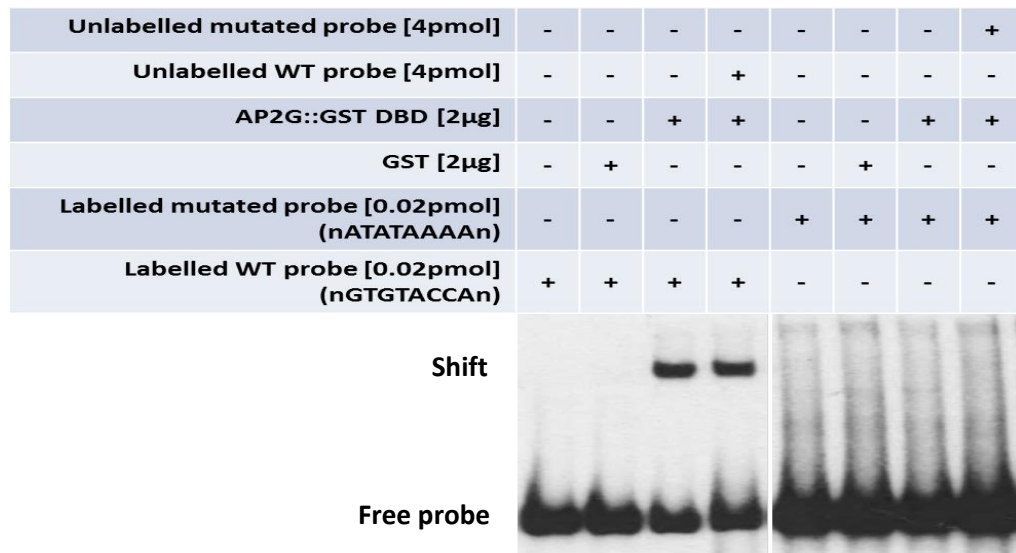
To complement this experiment, a second set of oligonucleotides was also commercially synthesised based on discovery of real/natural sequence bearing the core motif sequence (GTAC) in the immediate 1 kb of the 5' upstream region of the coding sequence of predicted target genes. The presence of the 5'upstream target sequence in the following genes was tested: 2 positions in PBANKA\_143750 (AP2-G; to explore its autoregulatory possibilities), and one position each in PBANKA\_103430 (AP2-G2; because it is thought to be downstream of AP2-G), and PBANKA\_081070 (SPM-1 or sub-pellicular microtubule protein-1; identified to be up-regulated in gametocytes by microarray analyses (Sinha et al., 2014; data not shown). The total length of the synthesised oligo in this case was ~60 bp with the target core motif in the centre of the sequence.

In both cases, the final oligos were synthesised each as 5'-biotinylated (labelled) and non-biotinylated (unlabelled) forms to serve as the target binding sequence and as a sequence-specific competitor, respectively. The labelled target oligos were used at 0.001 pmol/μl (total of 0.001 pmol per 20 μl reaction mix) and unlabelled target competitor oligos at 0.2 pmol/μl (a 200-fold excess to the labelled oligo and totalling to 0.2 pmol per 20 μl reaction mix). The details of the typical reaction mix are shown in Materials and Methods section 2.4.

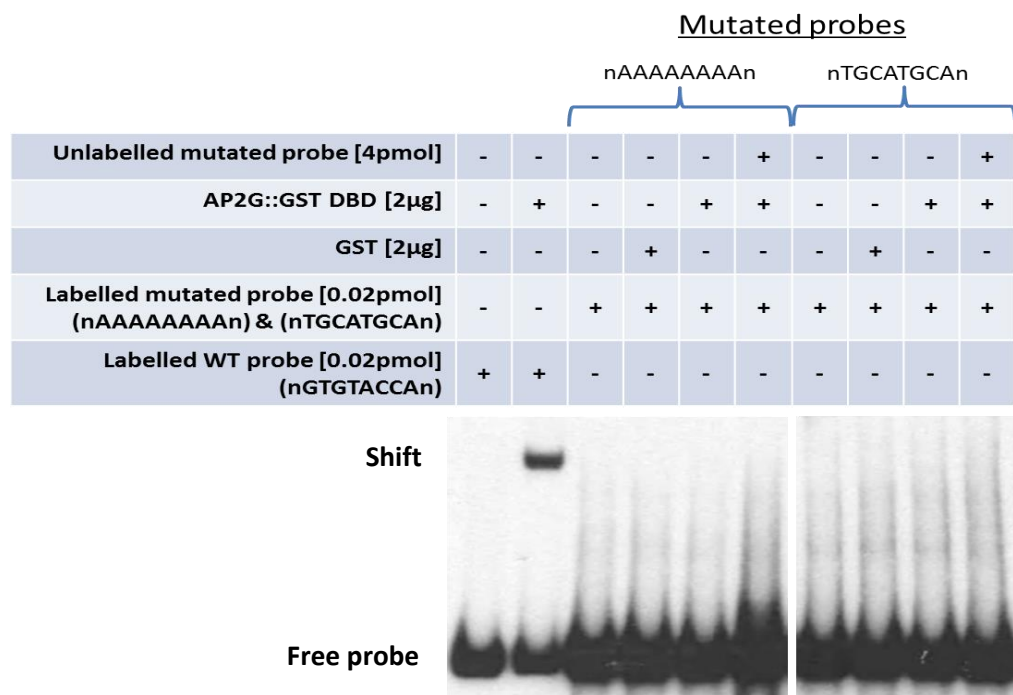
### 7.3.2 ***The DBD binds to the motif identified by PBM analysis in a sequence-specific manner***

The binding of the GST tagged DBD from *Pb* and *Pf* AP2-G was shown as a positive or upward shift of the bound labelled target oligonucleotide::DBD complex in the EMSAs (Figures 7-3 and 7-4). Figure 7-3 (a and b) shows the EMSA results using a synthesised labelled target sequence bearing the core (GTAC) recognition motif whereas Figure 7-4 shows the EMSA results using a naturally occurring *P. berghei* genomic sequence containing the predicted AP2-G target motif which was synthesised and labelled. The bound complex tends to migrate relatively slowly (as compared to the labelled target oligonucleotide alone) in a 12% PAGE gel under standard operating conditions as detailed in Materials and Methods, Section 2.4. This upward shift in the position of the band corresponding to the gel lane containing labelled oligo together with the GST tagged DBD was compared with the position of the corresponding bands in the gel lanes containing labelled oligonucleotide alone.

The binding of the GST tagged DBD to its predicted DNA recognition sequence was shown to be specific to the DBD as well as to the target sequence of the motif (Figures 7-3 and 7-4). This specificity was tested by conducting the EMSAs involving the GST tagged DBDs of *Pb* and *Pf* AP2-G and their target recognition motifs together with GST alone and mutated target recognition motif variants for AP2-G DBDs. The non-DBD protein used was GST alone. The mutated target recognition motif variants (Figure 7-3) included for the hypothetical labelled target included a poly-A variant (all 8 bases replaced with A), a GC/A variant (all G's and C's replaced with A) and a complementary base variant (complementary base change in the four core bases of the motif). Similarly, the mutated target recognition motif for the real/natural motif sequence included G to A and C to A base substitutions (Figure 7-4) and single nucleotide point mutation variants at various positions within the 4 bases included in the core target sequence. The mutated oligonucleotides were similar to the non-mutated ones outside the core recognition sequence and were synthesised, labelled and purified in exactly the same way as the non-mutated ones.



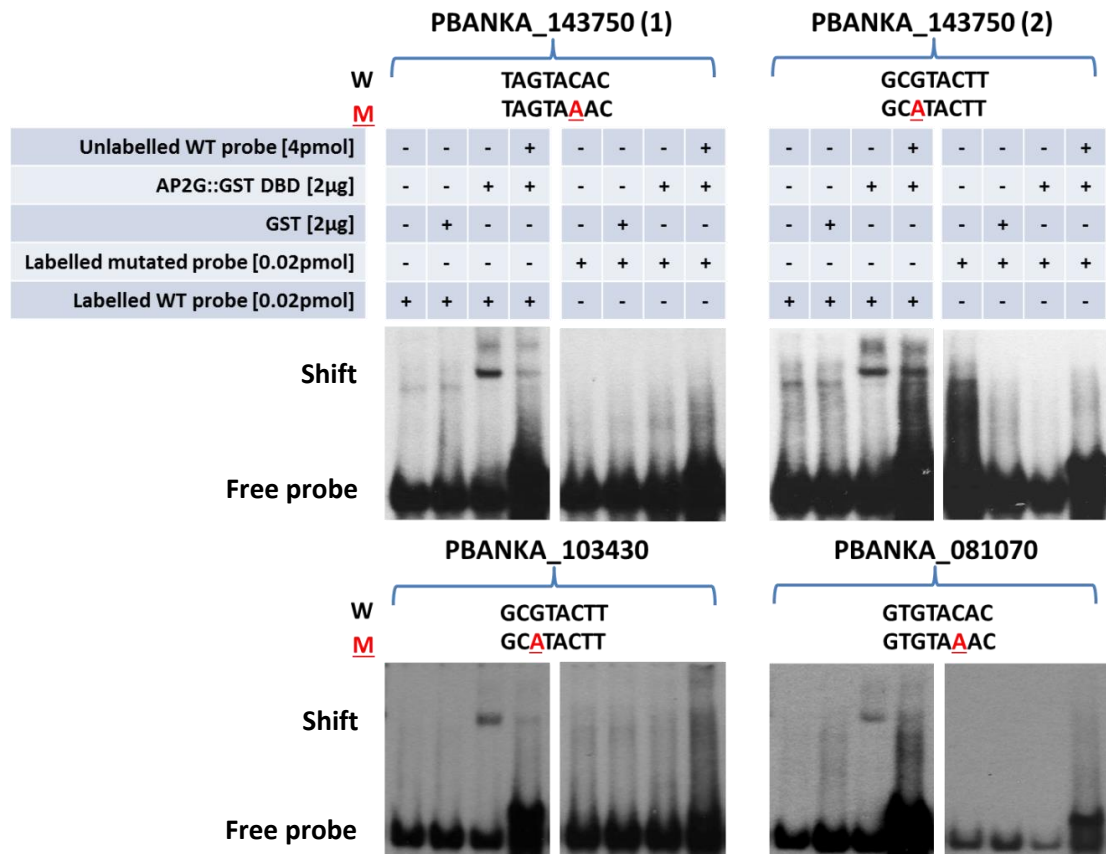
(a)



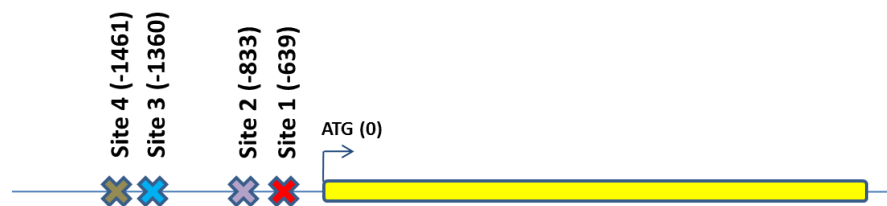
(b)

**Figure 7-3. *In vitro* demonstration of AP2-G DBD binding to its predicted synthetic recognition sequence:** Electrophoretic Mobility Shift Assay (EMSA) showing the recognition and subsequent binding of the GST tagged DBDs of *Pb* and *Pf* AP2-Gs to a hypothetical synthesised DNA sequence containing the predicted AP2-G DBD recognition motif in a sequence-specific manner. Figure 7-3 (a) shows the successful binding of the DBD to the cognate motif as a shift in the position of the band of the DBD::labelled oligonucleotide complex as compared to the baseline band position of the free labelled probe. It also shows that the binding (shift) is not affected in presence of a 400-fold excess of the unlabelled probe showing that the competitor effect was not obtained. The sequence specificity of the interaction between the DBD and the labelled probe is shown by absence of a shift when the labelled target sequence was changed from nGTGTACCAAn to nATATAAAAn. Further effect of a change in the target sequence is shown in Figure 7-3 (b) where the labelled target sequence was modified to nAAAAAAAAAn and nTGCATGCAn, re-establishing the sequence-specificity of the interactions.





(a)



**Site 1 (PBANKA\_081070; WT):** TTTCTTGTTTAGTGTTATAATAACGTGTACACACAACACATAAAAAATGCGCACCTATAT  
**Site 1 (PBANKA\_081070; Mut):** TTTCTTGTTTAGTGTTATAATAACGTGTAAACACAACACATAAAAAATGCGCACCTATAT

**Site 2 (PBANKA\_103430; WT):** ATATTTTAATCTGGTTCCTTTTCTGCGTACTTATCCAGAATGAATGTTTTAATTCATT  
**Site 2 (PBANKA\_103430; Mut):** ATATTTTAATCTGGTTCCTTTTCTGCGATACTTATCCAGAATGAATGTTTTAATTCATT

**Site 3 (PBANKA\_143750; WT):** TATGTTAAACATTGGCAAATTATAGTACACCCTAATAGTATCATTCTATTGCATCATT  
**Site 3 (PBANKA\_143750; Mut):** TATGTTAAACATTGGCAAATTATAGTAAACCCTAATAGTATCATTCTATTGCATCATT

**Site 4 (PBANKA\_143750; WT):** TAATATACTGTATGTATATATGCATGCGTACTTATGCATTATAAATTGTCTACACAATTT  
**Site 4 (PBANKA\_143750; Mut):** TAATATACTGTATGTATATATGCATGCGATACTTATGCATTATAAATTGTCTACACAATTT

(b)

**Figure 7-4. *In vitro* demonstration of AP2-G DBD binding to its predicted natural recognition sequence:** Shown are the interactions between GST tagged *Pb*AP2-G with synthesised oligonucleotide sequences which are found to occur naturally within the promoter region of the following predicted target genes for AP2-G based on the presence of occurrence of the target sequence GTAC within the immediate 1 kb of the promoter region of the following genes: two motifs in PBANKA\_143750 or AP2-G (1 and 2); one motif each in PBANKA\_103420 or AP2-G2 and PBANKA\_081070. The sequence-specific nature of the interaction is evident by the presence of band-shift (Figure 7-4 a) with the unmutated or WT probe (W) and its absence with the mutation variants of the probe (M; the mutated base is depicted as underlined). Successful competition of

the interaction is independently shown with addition of a 200-fold excess of the unlabelled probe. Figure 7-4 (b) shows a cartoon of the positional occurrence of the actual motif within the immediate 2 kb of the promoter region with reference to the translation start site (position 0) of the respective coding sequence of the gene (yellow bar). The relative distance of the individual motif sites (coloured cross) are also shown. The sequence used for the synthesis of the original wild type (WT) and the mutated (Mut) probe for each site is also depicted with the core motif shown in red and the mutated base is underlined.

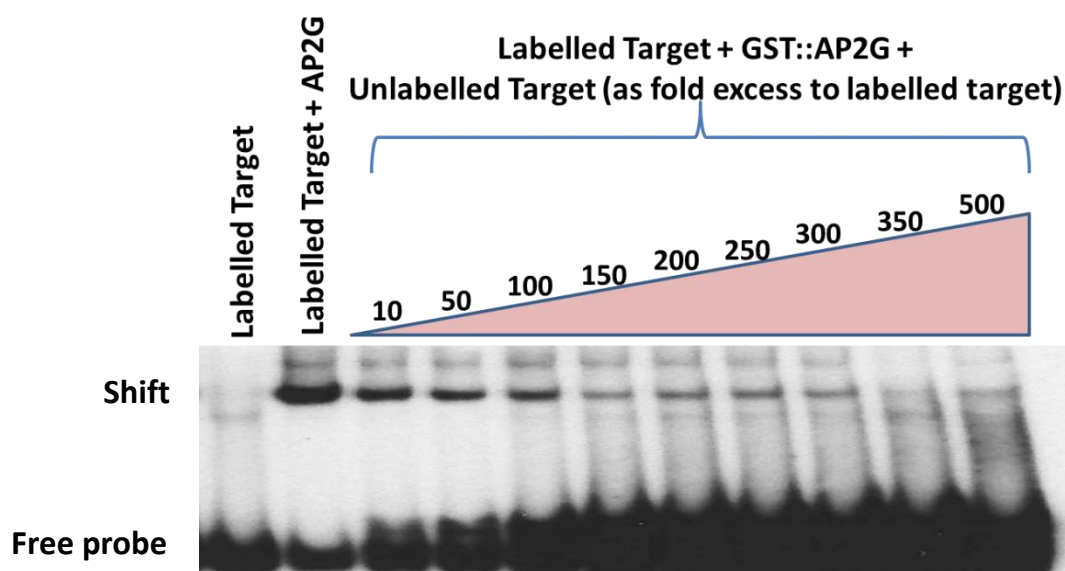
The presence of an upward shift in the band positions involving the correct/complementary combination pair of the DBD and the recognition sequence and the absence of any noticeable shift in the mismatched pair DBD and recognition sequence strongly suggests that the DBDs recognize and thus bind to a target DNA sequence in a sequence-specific manner, that is, the DBDs recognize and bind to a particular 4-base sequence (GTAC, in case of AP2-G) and any substitution, even a SNP, in the target sequence abolishes the binding.

### **7.3.3      *The binding is relieved by a sequence-specific competitor***

The binding (and thus the upward shift of the band) of the GST tagged DBD to its labelled cognate DNA sequence was non-evident in presence of excess unlabelled sequence-specific competitor in a dose-dependent manner. The competitor was the unlabelled counterpart of the target sequence oligonucleotide used at 200-fold excess of the labelled target sequence. The samples (DBD with labelled target oligonucleotide and DBD with unlabelled target oligonucleotide) were fractionated on a 12% PAGE gel under similar running conditions. The complex of the DBD with 200-fold excess of the unlabelled target oligonucleotide did not show a positive or upward shift in the gel as compared to the free labelled target probe. Together with a positive shift of the DBD + labelled target oligonucleotide and an absence of any shift with 200-fold excess of unlabelled target oligonucleotide, it was highly suggestive that an excess of the unlabelled target oligonucleotide was preferentially consumed in the formation of the complex with the DBD thus leaving no DBD available for making a complex with the labelled target oligonucleotide and thus no shift was observed. For reasons not known, the ability of excess unlabelled target oligonucleotide to prevent the binding of the DBD to the target motif

could only be demonstrated for the target oligonucleotides synthesised from a real 5'upstream promoter region of the AP2-G bearing the recognition motif (Figure 7-4a).

Based on the evidence of the band-shift being competed in presence of 200-fold excess of unlabelled target competitor, the relative amount of the unlabelled target oligonucleotide needed to make the gel-shift disappear was titrated (between the range 10 and 500-fold excess to the labelled probe). It was observed at ~350-fold excess of the unlabelled target oligonucleotide was sufficient to inhibit the appearance of the gel-shift (Figure 7-5).



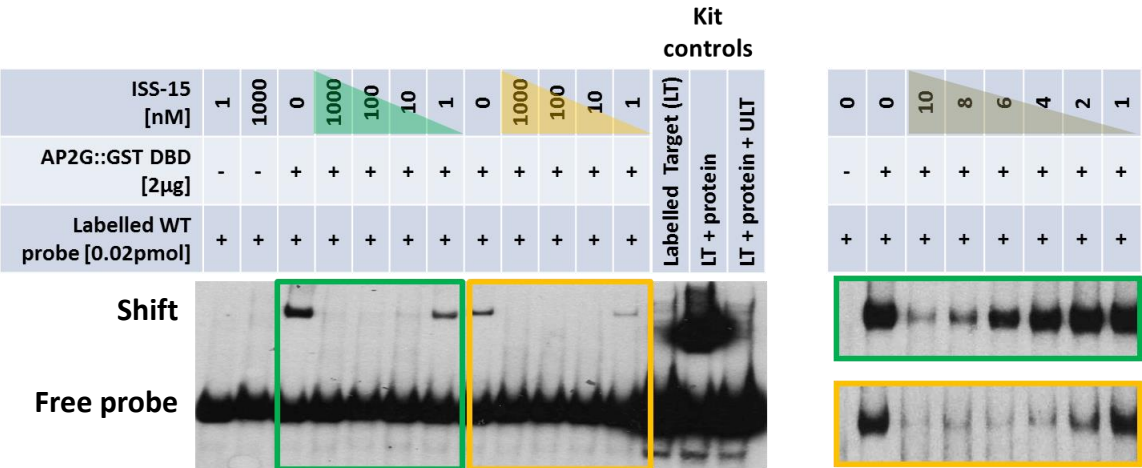
**Figure 7-5. EMSA showing dose-dependent titration of the binding competition between the GST tagged DBD of AP2-G with 0.01 pmol of the target labelled oligonucleotide and an increasing fold-excess concentration of the competitor probe i.e., unlabelled target oligonucleotide.** The range of concentration of the competitor was 10-fold to 500-fold excess as compared to the labelled counterpart, based on the blocking concentration of the unlabelled competitor from previous experiment (see Figure 7-4(a)). The end point was the concentration where the band-shift disappears which in this experiment was a 350-fold excess of the unlabelled target probe.

## 7.4 Polyamides and competition for DNA binding

Polyamides are a class of pyrrole-imidazole (Py-Im) synthetic chemical compounds that can be custom-designed to bind to DNA with high affinity and

sequence-specificity (Dervan, 2001; Dervan and Edelson, 2003). They have been shown to competitively recognize and subsequently bind to the cognate DNA sequence thus preventing the DBD from binding to the same site (Zhang et al, 2011; Manuel Llinas, Princeton University, unpublished data). Polyamides targeting specific DNA sequences of AP2-G with a higher affinity and specificity (in targeting the cognate DNA sequence) than the natural DBD can now be synthesised with high yield, cost-effectiveness and versatility (Glenn Burley and Ishwar Singh, University of Strathclyde, personal communication). Further, structural modifications can be made to provide better solubility in aqueous media (Matt Gregory, Hope College, USA, unpublished data; Glenn Burley and Ishwar Singh, University of Strathclyde, personal communication) so that these compounds can be tested both *in vitro* and *in vivo*.

To assess the role of these polyamides as potential competitors of the AP2-G DBD for the same target sequence *in vitro*, specific pyrrole- and imidazole- containing analogues of *P. berghei* AP2-G DBD (ISS-15) and the distinct AP2, *P. falciparum* AP2O DBD (ISS-33) were synthesised and kindly gifted to me by Ishwar Singh and Glenn Burley (University of Strathclyde, Glasgow). These polyamides had excellent binding affinity (K) and sequence specificity. ISS-15 was diluted to sub-nanomolar concentrations for performing competition assays using EMSAs. To determine the specific concentration of the polyamides necessary to preferentially bind to the target sequence, reflected as the point of disappearance of the band shift in the EMSA, a titration experiment was carried out with *P. berghei* GST::AP2-G DBD (green in Figure 7-6) and *P. falciparum* GST::AP2-G DBD (yellow in Figure 7-6) and ISS-15 concentration ranges between 1 nM and 1000 nM and which were then refined to between 1 nM and 10 nM. As these polyamides are low molecular weight entities when compared to the actual GST tagged DBD, the binding of the polyamides to the labelled cognate DNA sequence does not produce a shift hence disappearance of the band shift on adding excess ISS-15 is the read out (Figure 7-6). ISS-33 was also tested separately to competitively inhibit the AP2-G DBD binding with its recognition motif but expectedly, it did not block the shift (data not shown). This suggests that the synthesised polyamides are highly sequence specific so that ISS-15 will only recognize the AP2-G motif and there is no cross-recognition of the AP2-G motif by ISS-33 (which is specific for recognition of the AP2-O motif).



**Figure 7-6. EMSAs showing competitive interference of synthetic polyamide (ISS-15) on the interaction between the DBD of *Pb* AP2-G (green) and *Pf* AP2-G (yellow).** The amount of ISS-15 necessary to abolish the interaction between the AP2-G DBD and the target motif was titrated, the range being from 1 nM to 1000 nM (left panel). The optimum amount needed to prevent the interaction was less than 10 nM which was further titrated between 1 nM and 10 nM (right panel). It can be deduced that *in vitro*, 10 nM and 4-6 nM of ISS-15 sufficiently prevents almost all the interaction between the DBD and the target recognition motif in *Pb* and *Pf*, respectively.

## 7.5 Summary and discussion

Transcription factors play a very important role in controlling gene expression in eukaryotes. The paucity of established TFs in *Plasmodium* makes them an important candidate target for future malaria control strategies. The ApiAP2 class of TFs have already been proven to play important roles during ookinete and sporozoite stages of the *Plasmodium* life cycle in the invertebrate host. The discovery and subsequent characterization of AP2-G during this work, in terms of its capacity to influence multiple downstream transcriptional targets, through the demonstrated interactions between the DNA Binding Domain and the cognate target motifs indicates a critical role in controlling malaria transmission. AP2-G turns out to be a key regulator of the commitment of the asexual parasite to the sexual stage, and may be established as the master regulator of a major developmental switch in the life cycle of *Plasmodium*.

In this chapter, through EMSAs it was successfully demonstrated *in vitro* that the GST tagged DBD of AP2-G recognizes and interacts with a conserved *cis*-element sequence, nGTACn present within the promoter region of key regulatory genes - seen in the EMSAs as a positive band-shift. It was also shown

that this interaction is highly sequence specific, as even a single base mutation within the GTAC sequence abolishes the interaction. The sequence-specificity of the interaction was again demonstrated by the fact that a 200 to 350-fold excess of the unlabelled recognition sequence was able to compete out with the interaction. Further, the use of a custom-designed synthetic polyamide (ISS-15) that showed preferential binding to the recognition motif, thus preventing the actual binding of the AP2-G DBD, established the proof of concept that such compounds have the potential to be explored as future transmission-blocking candidates particularly with the current dearth of novel therapeutic and protective compounds for malaria control.

Further research questions stemming from this part of the work include concerns regarding the:

- functional role of such motifs *in vivo*, as the mere presence of a motif does not guarantee a related function
- *in vivo* testing through reporter constructs to check the binding of the DBD to the cognate DNA sequence present in the promoters/5'UTRs of the predicted genes
- minimum number of bases in the recognition target motif that are required for efficient binding
- nature of flanking sequences around the conserved recognition motif and its influence on binding
- effect of an interruption of the promoter in terms of increased distance of the target motif to the ATG
- role of the presence of one or more than one motifs in the promoter region
- attempts to predict the presence of these recognition motifs in the promoter and/or 5' UTRs of the critical sex-specific genes
- presence and role (auto-regulatory) of the target motif for AP2-G within the 2 kb upstream region of the coding sequence of AP2-G itself
- role of ISS polyamide compounds in preventing malaria transmission.

## **8 General discussion**

## 8 General discussion

Despite the renewed surge in enthusiasm and optimism regarding the attainability of the mammoth goal of eradicating malaria globally, the current armamentarium of anti-malaria measures offers little in the way of novel strategies. The appearance of foci of resistance to the present day frontline anti-malaria drugs, the Artemisinin Combination Therapy, in Thailand and the incipient risk of spreading the problem worldwide has generated serious concerns over the efficiency of malaria control and elimination (World Malaria Report, 2013). With no new drugs in the immediate pipeline and unconvincing results of Phase 3 trials of the much awaited malaria vaccine, RTS,S/AS01 (Agnandji et al., 2012; Moorthy et al., 2013), the urgent need to discover and develop newer modalities of treating and preventing transmission of malaria cannot be overemphasized (World Malaria Report, 2013). Specifically for malaria elimination and eradication, the identification of compounds which are effective in blocking the transmission of the disease by interrupting the transition between asexual and sexual mode of the life cycle of *Plasmodium* is urgently required.

Gametocytogenesis in malaria parasites and the completion of the sexual cycle are critical bottlenecks in the life cycle of the parasite. This thesis was aimed at dissecting and understanding the hitherto unknown molecular mechanisms (Eksi et al., 2012; Ikadai et al., 2013) regulating the generation of gametocytes in *P. berghei*, the rodent malaria model.

A classical forward genetics screen was used to generate the gametocyte-less mutants. Simultaneously, ten isogenic parasite lines, all derived from the male (GFP) and female (RFP) fluorescent reporter parental line 820, were maintained in asexual propagation by repeated mechanical passage in TO mice for about a year. Gametocytemia was monitored each week using flow cytometry and the desired gametocyte non-producer (GNP) phenotype was finally ascertained by absence of transmission of the GNP lines through mosquitoes.



The gametocytemia declined with the number of mechanical passages in all the ten lines by the end of the passage period but reached zero in only three of the potential ten lines (GNP9, GNP7 and GNP8). Various lines have been made in the past to generate such GNP mutants in *P. berghei* and *P. falciparum* but these mutants were never subjected to an immediate high throughput whole genome sequencing (WGS) approach to identify the attributable mutations in the genome (Micks, 1947a; Shirley et al., 1990; Janse et al., 1992; Day et al., 1993; Alano et al., 1995). Although the rate of accumulation of spontaneous evolutionary mutations in *Plasmodium* have recently been estimated to be low ( $1.7 \times 10^{-9}$  per base pair per generation) and comparable to other eukaryotes (Bopp et al., 2013), this study provides enough power and minimal mutational noise to specifically identify the causal GNP-associated mutations (if any) by comparative mapping of the genome sequences between the *de novo* GNP mutants and their isogenic gametocyte producer parent. Similar approaches have been used to identify the candidate genetic mutations associated with drug resistance in *Plasmodium* (Hunt et al., 2010; Martinelli et al., 2011; Modrzynska et al., 2012; Bright and Winzeler, 2013; Miotto et al., 2013; Neafsey, 2013).

The comparative genomic sequencing study identified an ApiAP2 family transcription factor, PBANKA\_143750, as the only gene being uniquely mutated in all the three *de novo* GNP mutants generated in this study. In addition, WGS of other pre-existing GNP lines K173cl1 and 233 also showed that AP2-G was mutated in these two lines. The ApiAP2, a family of 27 transcription factor proteins in *Plasmodium*, is the only identified specific transcription factor family in apicomplexans to date (Balaji et al., 2005). ApiAP2's have already been shown to be involved in regulating other life-cycle stages of *Plasmodium* such as ookinetes (AP2-O; Yuda et al., 2009), sporozoites (AP2-Sp; Yuda et al., 2010) and liver stage sporozoites (AP2-L; Iwanaga et al., 2012). To differentiate PBANKA\_143750 from the other three ApiAP2 members with a stage-specific role, the gene has been named AP2-G to indicate its role in *Plasmodium* gametocytogenesis.

This discovery of AP2-G is quite different in terms of the implicated role of the gene in the life cycle of the parasite. Whereas the other three identified AP2's have a role in progression of the life cycle stage of the parasite, AP2-G is

implicated in regulating a critical developmental switch that would determine the transmissibility of the parasite in nature.

The seminal finding of this study strongly suggested *AP2-G* as the master regulator of gametocytogenesis. However, the kinetics of declining gametocytogenesis in the generated GNP lines in the study, particularly GNP line 8, suggests that a cascade of genes that might need to be mutated before a line becomes a GNP line. If this is the case, then the findings of this study strongly argue that PBANKA\_143750 or *AP2-G* is the first in a series of genes that will be attributed to the absolute GNP phenotype. There might be other genes or other mutations within the same gene involved in the cascade which then could be hypothesised (for now) to be present upstream to *AP2-G*, mutation(s) in each contributing to a proportional fall in the number of gametocytes generated over time.

A battery of reverse genetics (gene deletion) experiments supplemented the observations of the classical forward genetics experiments in demonstrating that the *AP2-G*(-) mutants did not produce gametocytes and hence were not transmissible to mosquitoes. The three *AP2-G* knockout length variants also proved that the region bearing the three *de novo* GNP mutations (which was upstream to the DBD) and the DBD bearing region itself were independently associated with absence of gametocyte production. However, it would be quite interesting to knockout the 264 bp region downstream to the *AP2-G* DBD, to show its functional significance in regulating gametocytogenesis as evidenced by the location of the mutation in the GNP line 233. However, evidence suggests that extra sequences both at N- and C-terminal of the DBD in addition to the domain itself are critical for binding to the DNA (Balaji et al., 2005; Campbell et al., 2010).

It is also worth mentioning here that systematic ApiAP2 gene deletion experiments in *P. berghei* led to the concurrent but independent discovery of yet another ApiAP2 transcription factor, PBANKA\_103430 or *AP2-G2*, which when perturbed, produced a significant marked decrease in the gametocyte production (but not complete abolition) with only a few female gametocytes

observed which obviously could not be transmitted through the mosquitoes (Sinha et al., 2014; data not shown).

A high degree of conservation of the AP2-G DBD among the sexually reproducing apicomplexans like *Plasmodium spp.*, *Theileria spp.*, and *Babesia spp.*, signifies the importance of the protein as a critical regulator of sexual differentiation. Sufficient evidence is now available to believe that AP2-G is in fact the master regulator of commitment to gametocytogenesis in malaria parasites. Some of the most crucial evidences include concurrent identification of a *P. falciparum* orthologue of AP2-G (PFL1085w) through an independent approach (Kafsack et al., 2014) as essential for gametocytogenesis.

Also, direct growth competition assays between AP2-G knockout, WT *P. berghei* and AP2-G2 knockout revealed that AP2-G knockout mutants overgrew both the WT and AP2-G2 knockout mutants and AP2-G2 knockout mutants overgrew the WT indicating that AP2-G mutants completely lacked gametocytes thus offering a growth advantage to the asexually dividing parasites (Sinha et al., 2014; data not shown). This establishes the central role of AP2-G at the point of commitment to gametocytogenesis whereas AP2-G2 is involved downstream of AP2-G once the parasites are committed to form gametocytes. The role of AP2-G transcription factor as master regulator of the process of commitment to gametocytogenesis is further supported by the fact that AP2-G2 was never associated with any of the *de novo* GNP lines generated in the study reinforcing the hypotheses that AP2-G2 lies downstream to AP2-G. Further, based on the observation that once committed to sexual differentiation, the committed merozoites are either all females or all males (Silverstrini et al., 2000; Smith et al., 2000), it would be intriguing to know if there exists two flavours of AP2-G - AP2-G(m) and AP2-G(f)? Also, the answer to the question whether sexual stage commitment is accompanied by a concomitant sexual differentiation or is there a temporal gap between the two processes is of great interest.

However, as the data accumulates (Kafsack et al., 2014; Sinha et al., 2014), the regulation of the switch appears to be complex. It has been shown that all the AP2 transcription factors have at least one DNA binding domain

(DBD; conserved throughout the orthologues in *apicomplexa*) which recognizes and binds to a conserved sequence motif in a sequence-specific manner (Campbell et al., 2010; Painter et al., 2011). For PFL1085w, the *P. falciparum* orthologue of AP2-G, the motif, nnGTACnn, was discovered by PBMs and later shown to effectively bind the AP2 protein *in vitro* by EMSAs was also believed to be conserved. The present study (by using EMSAs) demonstrates that the nnGTACnn motif is in fact conserved for GST-tagged DBD of AP2-G identification and binding *in vitro*, and any single base substitution abrogates the binding of the AP2-G DBD with the cognate motif thus confirming the prediction made by Campbell et al (2010). The presence of the recognition motif in the upstream region of the *AP2-G* itself shows yet another level of control over transcription - a possible auto-regulation of *AP2-G* expression.

Identification of the downstream target genes for AP2-G would be the next immediate experimental task although preliminary reporter-based *in vivo* experiments conducted by Katie Hughes (Andy Waters' Lab) involving the promoters of some of the genes bearing the AP2-G recognition motif suggest that not all genes bearing the motif are under direct control of AP2-G. Though some of the motifs were shown to recognize and bind the GST-tagged AP2-G DBD *in vitro* using EMSAs, it might also be possible that DNA methylation and histone modifications at the nuclear periphery (Lopez-Rubio et al., 2009; Flueck et al., 2010) are involved, such that the mere presence of the recognition motif does not guarantee a regulatory control by AP2-G, thus offering fine tuning of gene expression in *Plasmodium* (Robertson et al., 2007; Campbell et al., 2010). Indeed evidence was obtained that protein-protein interactions involving the portion of *AP2-G* sequence other than the DBD might be involved in the regulatory complex, as evidenced by the location of the mutation in the pre-existing GNP line 233 (downstream to the DBD) in the final C-terminal 200aa of the >2300aa predicted protein.

Gene complementation efforts to restore the WT phenotype are required but not mandatory to prove a causal association if forward genetic studies are supplemented with coherent results from the reverse genetics counterpart. But, to prove the causality, complementation of the *de novo* GNP mutants was successfully attempted and the surviving WT parasites from the mixed

transfectant (WT+GNP) population were immediately rescued through mosquito passage to generate a clonal WT population after the bite-back experiments.

Complementation efforts in the study were challenged by at least two factors - first, the length of the gene to be complemented was ~7 kb thus giving rise to the inherent problems associated with cloning and manipulation of such a large and single DNA fragment; and second, the reversal of the GNP phenotype to a gametocyte-producer phenotype had to happen against a strong natural selection pressure in favour of the GNP parasites, posing problems in negatively selecting the transgenic complemented parasites.

A proof of principle for identifying and developing AP2-G as an anti-gametocyte agent was also established in the study by using a synthetic polyamide compound (ISS-15) designed by Ishwar Singh and Glenn Burley, University of Strathclyde, UK. Such polyamides offer molecular mimicry and preferentially block the DBD recognition motif thus preventing the formation of the transcription initiation complex (Zhang et al, 2011). ISS-15 was shown by EMSAs to preferentially bind to the AP2-G recognition motif and thus abolish the binding of the AP2-G DBD with the motif. The polyamides can be structurally customized to offer specificity of binding to any DNA sequence and thus can selectively target/inhibit the regulation of particular stage-specific transcription events.

Taken together, the collective evidence strongly favours the conclusion that PbAP2-G has a pivotal role in the commitment of asexually dividing malaria parasites, causing them to developmentally switch to the sexual mode of reproduction and generate the sexual precursors in the mammalian host - the gametocytes. The fact that such a transition occurs within the mammalian host offers a glimpse of hope towards identification of new anti-gametocyte targets for vaccine/drugs based on the molecular basis of *AP2-G* functioning and its vulnerability to successful interruption, at least, *in vitro*. However, additional modes of action of AP2-G including possible protein-protein mediated actions and the epigenetic control of *AP2-G* gene expression offer possible intricacies and, at the same time, insight into the apparently simple molecular functioning of *PbAP2-G* and its regulation.

Our current understanding of the mechanisms of transcriptional regulation by the ApiAP2 TFs in general offers hope for exploiting them as a new generation of drug targets. The areas to explore, however, include the drug candidate's ability to modulate either the synthesis/stimulation of ApiAP2's, the regulation of their activity by ligands or phosphorylation events, their binding to the DNA, the protein-protein interactions mediated by the TFs and/or at the epigenetic level, by stimulating activity of transcription factors' silencers.

Blocking TF synthesis/activation: This approach is likely to be of use only where it is desirable to switch on or off an entire bank of genes regulated by a particular TF. In such cases, it would be more effective to target the synthesis of the factor itself. Alternatively, the molecules/pathways involved in the activation of a TF could be targeted. For example, a metabolomics-based approach to study gametocyte non-producers in details may reveal metabolic signatures of commitment to sexual development and such signature molecules can also be explored. In many situations however, it will be preferable to develop ways of modulating the particular activity(ies) of the transcription factor, particularly when downstream roles of the target genes remain unidentified.

Blocking DNA binding: Binding of the TF to specific binding site/s in the target DNA molecule is necessary for TF action and thus represents an obvious target for therapeutic drugs. Evidence from the specific role of the DBD of AP2-O clearly suggests that blocking of DBD-binding to the target DNA motif could be explored as a drug target. Further evidence generated in this research regarding the *in vitro* ability of synthetic custom-designed polyamides in blocking the binding of the TF DBD to the recognition motif in a sequence-specific manner is also supportive of the approach. The property of the TF to heterodimerize in order to be able to bind to the recognizable cis-element (Lindner et al., 2010) also could be exploited. Indeed, inhibition of E2F/DP-1 heterodimerization actually achieved the effect of inducing apoptosis in tumour cells by preventing DNA-binding of the TF which required heterodimer formation (Bandara, Girling, La Thangue, 1997). A potential drawback of blocking specific activity(ies) of TFs is the possibility of undiscovered / unexpected redundancy in action that bypasses the DBD/DNA interaction such as the presence of functional protein-

protein interactions or epigenetics based actions and these will tend to be overlooked.

Blocking protein-protein interactions: Although the precise role of the DBD is well established as a pre-requisite for the downstream action of the TF, it might be possible that areas of the TF gene other than the DBD might be implicated in causing the altered gene expression and phenotype. In such a context, it is essential that the possibility of non-DBD mediated actions of the TF, for example, protein-protein interaction based regulation should not be ruled out.

Some of the other practical challenges in developing a TF-based drug target include the need to accumulate and sustain the proposed drug within the parasites, presence of multifaceted cross-talk between different transcription factors and their interactions with target genes across various tissues, cellular contexts and temporal settings, and the higher order chromatin organization of the TFs affecting their localization, turnover and gene targeting potential (Karamouzis and Papavassiliou, 2011).

However, using transcription factor blockers as anti-malaria candidates offers certain advantages. There is no known human orthologue for the ApiAP2 TF family so this would mean that if a transmission-blocking candidate is designed for intended human intake, it may have limited interference with human physiology and hence better compliance. Secondly, in the absence of mosquito bites, it would be an evolutionary advantage for the parasites to support and sustain the interruption in gametocytogenesis (Carter et al., 2013) and if such GNP changes are irreversible, then that would be a dead end for the parasite to escape the mammalian host. Although theoretically sound and convincing, a lot would depend how this GNP phenotype would behave in nature particularly regarding the changes in transmission dynamics and pathogenicity of the GNP parasites, to name a few. Importantly, the possibility of an undiscovered redundancy in the regulation of the gametocytogenesis switch cannot be ruled out at this stage, particularly when the involvement of a cascade of genes is hypothesized as part of the regulatory machinery.

Despite providing a new knowledge base about the molecular switch implicated in regulating gametocytogenesis in malaria parasites, the study opens the doors to a plethora of research questions for the malaria fraternity to answer and further refine knowledge on gametocytogenesis. Some of the research challenges are briefly mentioned below.

The lines that did not end up as GNP but rather as low gametocyte producers could be fruitful in identifying the hypothesized cascade of genes (if any) positioned upstream to the *AP2-G* or any other molecular mechanisms responsible for the gradual lowering of gametocytemia in these lines. The application of mathematical modeling to predict the almost precise point in the gametocytemia kinetics which shows periods of significant decline of gametocyte numbers could be very useful to select the particular cryo-preserved stabilate for whole genome sequencing.

These predictions would then help us identify the series of mutations and investigate whether these mutations significantly bring a change in the kinetics and thus reduce the proportion of gametocytes until a critical mutation affects *AP2-G* and totally stops gametocyte production or alternatively, *AP2-G* is mutated in a stochastic process thus abolishing gametocyte generation. This would be particularly interesting to explore in the GNP line 8 which showed a gradual decline of gametocytemia throughout the passage period.

Similar observations were also recorded when a GNP line was successfully complemented. Unintended passages of the line prior to the mosquito transmission resulted in the loss of complemented parasites. It may well be reasoned that the complementation happened in the background of the already existing GNP parasites which gradually replaced the complemented counterparts in the absence of transmission. It would be worthwhile to show that this indeed happened by cloning the complemented GNP line and comparing the clones. Further, sequencing the *AP2-G* gene from other pre-existing GNP lines in the Andy Waters' lab (University of Glasgow) would throw further light on the variety of mutations (if any) that contribute to the GNP phenotype in these lines.



What are the precise signaling cues and mechanisms that turn the molecular switch on is yet another area of practical potential importance. Although recent literature suggests that the signals are mediated via microvesicles and exosome-like vesicles (Mantel et al., 2013; Regev-Rudzki et al, 2013), the precise nature and mediators of signaling are yet to be explored. In this context, the preserved frozen plasma from each week of the passage experiment could prove beneficial to decode the metabolomic signatures which can reveal the precise signaling mechanisms in operation at or immediately before the actual commitment to gametocytogenesis (delPortillo and Chitnis, 2013). To add to the complexity of the issue, it would be interesting to know that if the GNP lines are exposed to mosquitoes, will the same signalling mechanism which was thought to be providing the sensory cues for the parasite to abolish the generation of gametocytes in the absence of mosquitoes, become recruited to produce gametocytes and if this is there, a selective advantage over the GNP population? This would suggest that these genetic changes may not be irreversible, at least for this GNP phenotype in *Plasmodium*.

As the open reading frame of the gene does not possess any known functional domain other than the DBD, the functional role of the rest of the coding sequence remains unclear. Moreover, with the presence of a nonsense mutation located immediately downstream to the DBD in the pre-existing GNP line 233, the functional implications of the peri-DBD areas require investigation. Such a mutation is expected to result in a fully functional DBD which might indicate another level of gene expression regulation by mediation of protein-protein and/or protein-DNA interactions involving the region of the protein downstream of the mutation.

In addition to the rest of the open reading frame of *AP2-G* outside the DBD, the role of the un-translated regions (UTRs), particularly the 5'UTR of the gene, also warrant further investigation. It is worthwhile to note that when the 5'UTR of the gene was interfered with big insertion of the drug selectable marker cassette (SMC) while complementing a knockout line with the recombineering based repair construct (Chapter 6), the expression of the gene was negatively affected in terms of the ability of the knock out line to regenerate gametocytes. It remains unclear whether the loss of expression

occurred due to the DBD recognition motifs getting interrupted by the SMC or possibly nucleosome phasing getting out of phase because of the extra length of DNA. Whatever the reason may be, the *AP2-G* promoter was interrupted and rendered non-functional due to the incorporation of the SMC in the 5'UTR of the gene.

Further refining of the interaction between the AP2-G DBD and recognition motif would help answering the questions regarding uncertainty in the precise control of gametocytogenesis. These include, but not limited to, the number of motifs required for gene expression, the identification of the minimum number of bases required for efficient and specific binding, the distance between two motifs, the role of flanking sequences in binding, the distance of the motif relative to the ATG, the role of multiple DBDs and multiple recognition motifs even for a single DBD, the role of *AP2-G* coding sequence regions out with DBD, nucleosome phasing and many more.

## **Appendices**

**Table A1: Line 1 - weekly parasitemia & gametocytemia by flow cytometry** (G=Gams, M=Males, F=Females, A=asexuals) before discounting for control readings in uninfected mouse (UI). Discounted data is shown as T' (total parasitemia), G' (gams), M' (males) & F' (females)

Weeks	Undicounted Total (%)	Total in control (%)	Discounted (as total %)				Discounted Total (%)	Discounted (as % of total parasites)		
			G	M	F	A		G'	M'	F'
1	4.14	0	0.6	0.29	0.31	3.54	4.14	14.49	7.00	7.49
2	4.52	0	0.49	0.37	0.12	4.03	4.52	10.84	8.19	2.65
3	2.96	0	0.33	0.22	0.11	2.63	2.96	11.15	7.43	3.72
4	3.34	0	0.73	0.33	0.4	2.61	3.34	21.86	9.88	11.98
5	3.74	0	0.99	0.46	0.53	2.75	3.74	26.47	12.30	14.17
6	16.14	0	0.24	0.15	0.09	15.9	16.14	1.49	0.93	0.56
7	3.33	0	0.24	0.17	0.07	3.09	3.33	7.21	5.11	2.10
8	1.67	0	0.12	0.08	0.04	1.55	1.67	7.19	4.79	2.40
9	0.72	0	0.04	0.03	0.01	0.68	0.72	5.56	4.17	1.39
10	2.44	0	0.63	0.37	0.26	1.81	2.44	25.82	15.16	10.66
11	3.26	0	0.47	0.31	0.16	2.79	3.26	14.42	9.51	4.91
12	2.96	0	0.63	0.3	0.33	2.33	2.96	21.28	10.14	11.15
13	3.01	0	0.24	0.11	0.13	2.77	3.01	7.97	3.65	4.32
14	3.22	0	0.41	0.24	0.17	2.81	3.22	12.73	7.45	5.28
15	2.88	0	0.72	0.37	0.35	2.16	2.88	25.00	12.85	12.15
16	3.81	0	0.86	0.42	0.44	2.95	3.81	22.57	11.02	11.55
17	3.7	0	0.95	0.57	0.38	2.75	3.7	25.68	15.41	10.27
18	2.58	0	0.21	0.15	0.06	2.37	2.58	8.14	5.81	2.33
19	4.94	0	0.49	0.23	0.26	4.45	4.94	9.92	4.66	5.26
20	3.28	0	0.65	0.35	0.3	2.63	3.28	19.82	10.67	9.15
21	3.04	0	0.77	0.56	0.21	2.27	3.04	25.33	18.42	6.91
22	3.97	0	0.46	0.24	0.22	3.51	3.97	11.59	6.05	5.54
23	3.81	0	0.69	0.3	0.39	3.12	3.81	18.11	7.87	10.24
24	4.7	0	0.49	0.28	0.21	4.21	4.7	10.43	5.96	4.47
25	7.09	0	0.63	0.33	0.3	6.46	7.09	8.89	4.65	4.23
26	4.73	0	0.38	0.15	0.23	4.35	4.73	8.03	3.17	4.86
27	11.06	0	0.29	0.17	0.12	10.8	11.06	2.62	1.54	1.08
28	5.92	0	0.47	0.26	0.21	5.45	5.92	7.94	4.39	3.55
29	4.2	1.31	0.28	0.17	0.11	3.92	2.89	9.69	5.88	3.81
30	7.59	0.79	0.39	0.12	0.27	7.2	6.8	5.74	1.76	3.97
31	6.25	3.62	0.27	0.09	0.18	5.98	2.63	4.32	1.44	2.88
32	8.18	3.05	0.15	0.07	0.08	8.03	5.13	1.8337	0.86	0.98
33	4.14	0.2	0.32	0.12	0.2	3.82	3.94	7.7295	2.90	4.83
34	2.68	0.19	0.07	0.02	0.05	2.61	2.49	2.6119	0.75	1.87
35	3.44	0.27	0.23	0.08	0.15	3.21	3.17	6.686	2.33	4.36
36	5.61	0.7	0.14	0.1	0.04	5.47	4.91	2.4955	1.78	0.71
37	6.42	0.34	0.29	0.08	0.21	6.13	6.08	4.5171	1.25	3.27
38	12.27	6.74	0.22	0.16	0.06	12.1	5.53	1.793	1.30	0.49
39	7.38	2.69	0.16	0.07	0.09	7.22	4.69	2.168	0.95	1.22
40	10.76	5.18	0.49	0.21	0.28	10.3	5.58	4.5539	1.95	2.60
41	20.98	1.92	0.46	0.2	0.26	20.5	19.06	2.1926	0.95	1.24
42	9.75	2.95	0.36	0.1	0.26	9.39	6.8	3.6923	1.03	2.67
43	9.81	0.32	0.31	0.16	0.15	9.5	9.49	3.16	1.63	1.53
44	9.09	1.96	0.31	0.16	0.15	8.78	7.13	3.4103	1.76	1.65
45	10.06	1.83	0.19	0.11	0.08	9.87	8.23	1.8887	1.09	0.80
46	12.26	3.65	0.19	0.11	0.08	12.1	8.61	1.5498	0.90	0.65
47	12.12	2.59	0.29	0.13	0.16	11.8	9.53	2.3927	1.07	1.32
48	7.69	1.12	0.46	0.12	0.34	7.23	6.57	5.9818	1.56	4.42
49	8.39	1.75	0.19	0.04	0.15	8.2	6.64	2.2646	0.48	1.79
50	12.28	1.76	1.02	0.81	0.21	11.3	10.52	8.3062	6.60	1.71
51	7.91	0.9	0.16	0.04	0.12	7.75	7.01	2.0228	0.51	1.52
52	9.98	1.11	0.14	0.06	0.08	9.84	8.87	1.4028	0.60	0.80

**Table A2: Line 2 - weekly parasitemia & gametocytemia by flow cytometry** (G=Gams, M=Males, F=Females, A=asexuals) before discounting for control readings in uninfected mouse (UI). Discounted data is shown as T' (total parasitemia), G' (gams), M' (males) & F' (females)

Weeks	Undiscounted Total (%)	Total in control (%)	Discounted (as total %)				Discounted Total (%)	Discounted (as % of total parasites)		
			G	M	F	A		G'	M'	F'
1	0.21		0.03	0.02	0.01	0.18	0.21	14.29	9.52	4.76
2	4.03		0.67	0.32	0.35	3.36	4.03	16.63	7.94	8.68
3	0.6		0.04	0.03	0.01	0.56	0.6	6.67	5.00	1.67
4	1.28		0.1	0.07	0.03	1.18	1.28	7.81	5.47	2.34
5	1.98		0.19	0.12	0.07	1.79	1.98	9.60	6.06	3.54
6	2.42		0.17	0.12	0.05	2.25	2.42	7.02	4.96	2.07
7	6.87		0.9	0.39	0.51	5.97	6.87	13.10	5.68	7.42
8	7.55		0.63	0.32	0.31	6.92	7.55	8.34	4.24	4.11
9	3.91		0.93	0.36	0.57	2.98	3.91	23.79	9.21	14.58
10	8.75		0.82	0.36	0.46	7.93	8.75	9.37	4.11	5.26
11	5.08		0.49	0.24	0.25	4.59	5.08	9.65	4.72	4.92
12	2.73		0.23	0.15	0.08	2.5	2.73	8.42	5.49	2.93
13	7.32		1.45	0.86	0.59	5.87	7.32	19.81	11.75	8.06
14	4.27		0.4	0.33	0.07	3.87	4.27	9.37	7.73	1.64
15	4.99		0.8	0.59	0.21	4.19	4.99	16.03	11.82	4.21
16	4.59		0.38	0.2	0.18	4.21	4.59	8.28	4.36	3.92
17	7.72		1.28	0.55	0.73	6.44	7.72	16.58	7.12	9.46
18	5.14		0.37	0.25	0.12	4.77	5.14	7.20	4.86	2.33
19	7.06		0.63	0.41	0.22	6.43	7.06	8.92	5.81	3.12
20	5.85		0.34	0.22	0.12	5.51	5.85	5.81	3.76	2.05
21	13.79		0.61	0.24	0.37	13.2	13.79	4.42	1.74	2.68
22	6.58		0.51	0.33	0.18	6.07	6.58	7.75	5.02	2.74
23	4.49	1.77	0.38	0.21	0.17	4.11	2.72	13.97	7.72	6.25
24	6.02	0.89	0.34	0.23	0.11	5.68	5.13	6.63	4.48	2.14
25	6.38	3.62	0.34	0.16	0.18	6.04	2.76	12.32	5.80	6.52
26	8.47	3.05	0.09	0.05	0.04	8.38	5.42	1.66	0.92	0.74
27	5.65	0.24	0.48	0.35	0.13	5.17	5.41	8.87	6.47	2.40
28	5.05	0.19	0.11	0.08	0.03	4.94	4.86	2.26	1.65	0.62
29	5.52	0.27	0.5	0.24	0.26	5.02	5.25	9.52	4.57	4.95
30	5.08	0.38	0.51	0.32	0.19	4.57	4.7	10.85	6.81	4.04
31	4.48	0.34	0.24	0.06	0.18	4.24	4.14	5.36	1.339	4.0179
32	18.12	10.04	0.51	0.21	0.3	17.6	8.08	2.81	1.159	1.6556
33	7.9	2.69	0.34	0.13	0.21	7.56	5.21	4.3	1.646	2.6582
34	9.94	5.18	0.45	0.18	0.27	9.49	4.76	4.53	1.811	2.7163
35	11.08	1.92	0.83	0.31	0.52	10.3	9.16	7.49	2.798	4.6931
36	10.26	2.95	0.23	0.06	0.17	10	7.31	2.24	0.585	1.6569
37	7.58	0.32	0.38	0.26	0.12	7.2	7.26	5.01	3.43	1.5831
38	20.38	16.99	0.46	0.13	0.33	19.9	3.39	2.26	0.638	1.6192
39	9.48	1.83	0.17	0.11	0.06	9.31	7.65	1.79	1.16	0.6329
40	13.95	3.65	0.27	0.19	0.08	13.7	10.3	1.94	1.362	0.5735
41	12.86	2.59	0.43	0.22	0.21	12.4	10.27	3.34	1.711	1.633
42	8.31	1.12	0.66	0.35	0.31	7.65	7.19	7.94	4.212	3.7304
43	7.49	1.75	0.19	0.13	0.06	7.3	5.74	2.54	1.736	0.8011
44	11.93	1.76	1.5	0.99	0.51	10.4	10.17	12.6	8.298	4.2749
45	8.1	2.92	0.1	0.06	0.04	8	5.18	1.23	0.741	0.4938
46	7.63	1.74	0.6	0.22	0.38	7.03	5.89	7.86	2.883	4.9803
47	10.61	2.33	0.25	0.13	0.12	10.4	8.28	2.36	1.225	1.131
48	6.27	0.2	0.13	0.04	0.09	6.14	6.07	2.07	0.638	1.4354

**Table A3: Line 3 - weekly parasitemia & gametocytemia by flow cytometry** (G=Gams, M=Males, F=Females, A=asexuals) before discounting for control readings in uninfected mouse (UI). Discounted data is shown as T' (total parasitemia), G' (gams), M' (males) & F' (females)

Weeks	Undiscounted Total (%)	Total in control (%)	Discounted (as total %)				Discounted Total (%)	Discounted (as % of total parasites)		
			G	M	F	A		G'	M'	F'
1	3.15		0.29	0.19	0.1	2.86	3.15	9.21	6.03	3.17
2	9.67		0.71	0.6	0.11	8.96	9.67	7.34	6.20	1.14
3	3.44		0.42	0.28	0.14	3.02	3.44	12.21	8.14	4.07
4	4.03		0.73	0.43	0.3	3.3	4.03	18.11	10.67	7.44
5	3.58		0.62	0.27	0.35	2.96	3.58	17.32	7.54	9.78
6	1.69		0.14	0.11	0.03	1.55	1.69	8.28	6.51	1.78
7	2.9		0.21	0.14	0.07	2.69	2.9	7.24	4.83	2.41
8	4.34		0.4	0.29	0.11	3.94	4.34	9.22	6.68	2.53
9	0.92		0.06	0.04	0.02	0.86	0.92	6.52	4.35	2.17
10	0.36		0.03	0.02	0.01	0.33	0.36	8.33	5.56	2.78
11	3.65		0.34	0.24	0.1	3.31	3.65	9.32	6.58	2.74
12	2.77		0.21	0.14	0.07	2.56	2.77	7.58	5.05	2.53
13	6.73		0.48	0.2	0.28	6.25	6.73	7.13	2.97	4.16
14	2.81		0.73	0.28	0.45	2.08	2.81	25.98	9.96	16.01
15	4.07		0.38	0.21	0.17	3.69	4.07	9.34	5.16	4.18
16	3.21		0.55	0.33	0.22	2.66	3.21	17.13	10.28	6.85
17	3.02		0.36	0.19	0.17	2.66	3.02	11.92	6.29	5.63
18	1.67		0.11	0.08	0.03	1.56	1.67	6.59	4.79	1.80
19	4.15		0.49	0.26	0.23	3.66	4.15	11.81	6.27	5.54
20	3.49		0.49	0.33	0.16	3	3.49	14.04	9.46	4.58
21	5.08		0.55	0.39	0.16	4.53	5.08	10.83	7.68	3.15
22	3.48		0.53	0.31	0.22	2.95	3.48	15.23	8.91	6.32
23	4.66		0.59	0.29	0.3	4.07	4.66	12.66	6.22	6.44
24	3.33		0.45	0.17	0.28	2.88	3.33	13.51	5.11	8.41
25	3.5		0.56	0.37	0.19	2.94	3.5	16.00	10.57	5.43
26	8.79		0.39	0.17	0.22	8.4	8.79	4.44	1.93	2.50
27	7.45		0.55	0.22	0.33	6.9	7.45	7.38	2.95	4.43
28	12.69		0.37	0.18	0.19	12.32	12.69	2.92	1.42	1.50
29	6.42		0.38	0.26	0.12	6.04	6.42	5.92	4.05	1.87
30	4.98	1.77	0.19	0.09	0.1	4.79	3.21	5.92	2.80	3.12
31	7.06	0.79	0.36	0.18	0.18	6.7	6.27	5.74	2.87	2.87
32	6.4	3.62	0.08	0.05	0.03	6.32	2.78	2.88	1.80	1.08
33	5.91	3.05	0.23	0.11	0.12	5.68	2.86	3.89	1.86	2.03
34	4.42	0.24	0.25	0.14	0.11	4.17	4.18	5.66	3.17	2.49
35	8.8	0.19	0.11	0.07	0.04	8.69	8.61	1.25	0.80	0.45
36	7.97	0.37	0.22	0.14	0.08	7.75	7.6	2.76	1.76	1.00
37	8.89	0.7	0.43	0.27	0.16	8.46	8.19	4.84	3.04	1.80
38	4.76	0.34	0.1	0.07	0.03	4.66	4.42	2.10	1.47	0.63
39	15.87	6.74	0.19	0.11	0.08	15.68	9.13	1.20	0.69	0.50
40	10.61	2.08	0.33	0.19	0.14	10.28	8.53	3.11	1.79	1.32
41	14.25	5.18	0.27	0.09	0.18	13.98	9.07	1.89	0.63	1.26
42	19.04	1.92	0.34	0.13	0.21	18.7	17.12	1.79	0.68	1.10
43	14.17	2.95	0.47	0.14	0.33	13.7	11.22	3.32	0.99	2.33
44	11.11	0.32	0.3	0.19	0.11	10.81	10.79	2.70	1.71	0.99
45	24.03	16.99	0.32	0.1	0.22	23.71	7.04	1.33	0.42	0.92
46	11.41	1.83	0.33	0.23	0.1	11.08	9.58	2.89	2.02	0.88
47	14.08	3.65	0.35	0.21	0.14	13.73	10.43	2.49	1.49	0.99
48	8.81	2.59	0.23	0.17	0.06	8.58	6.22	2.61	1.93	0.68
49	6.64	1.12	0.16	0.06	0.1	6.48	5.52	2.41	0.90	1.51
50	8.58	1.75	0.14	0.05	0.09	8.44	6.83	1.63	0.58	1.05
51	11.04	0.4	0.39	0.25	0.14	10.65	10.64	3.53	2.26	1.27
52	8.48	2.92	0.18	0.12	0.06	8.3	5.56	2.12	1.42	0.71
53	10.63	1.74	0.5	0.16	0.34	10.13	8.89	4.70	1.51	3.20

**Table A4: Line 4 - weekly parasitemia & gametocytemia by flow cytometry** (G=Gams, M=Males, F=Females, A=asexuals) before discounting for control readings in uninfected mouse (UI). Discounted data is shown as T' (total parasitemia), G' (gams), M' (males) & F' (females)

Weeks	Undiscounted Total (%)	Total in control	Discounted (as total %)				Discounted Total (%)	Discounted (as % of total parasites)		
			G	M	F	A		G'	M'	F'
1	4.08		0.56	0.34	0.22	3.52	4.08	13.73	8.33	5.39
2	6.94		1.44	1.16	0.28	5.5	6.94	20.75	16.71	4.03
3	3.64		0.16	0.1	0.06	3.48	3.64	4.40	2.75	1.65
4	4.67		0.52	0.27	0.25	4.15	4.67	11.13	5.78	5.35
5	2.18		0.49	0.25	0.24	1.69	2.18	22.48	11.47	11.01
6	4.24		0.48	0.33	0.15	3.76	4.24	11.32	7.78	3.54
7	5.05		0.44	0.24	0.2	4.61	5.05	8.71	4.75	3.96
8	4.72		0.42	0.28	0.14	4.3	4.72	8.90	5.93	2.97
9	0.48		0.03	0.02	0.01	0.45	0.48	6.25	4.17	2.08
10	0.89		0.03	0.02	0.01	0.86	0.89	3.37	2.25	1.12
11	0.99		0.04	0.03	0.01	0.95	0.99	4.04	3.03	1.01
12	2.67		0.19	0.14	0.05	2.48	2.67	7.12	5.24	1.87
13	3.99		0.27	0.16	0.11	3.72	3.99	6.77	4.01	2.76
14	4.08		0.49	0.23	0.26	3.59	4.08	12.01	5.64	6.37
15	4.25		0.2	0.1	0.1	4.05	4.25	4.71	2.35	2.35
16	3.58		0.46	0.33	0.13	3.12	3.58	12.85	9.22	3.63
17	2.95		0.37	0.19	0.18	2.58	2.95	12.54	6.44	6.10
18	3.06		0.46	0.29	0.17	2.6	3.06	15.03	9.48	5.56
19	3.91		0.71	0.42	0.29	3.2	3.91	18.16	10.74	7.42
20	3.4		0.27	0.2	0.07	3.13	3.4	7.94	5.88	2.06
21	3.57		0.57	0.42	0.15	3	3.57	15.97	11.76	4.20
22	2.43		0.29	0.21	0.08	2.14	2.43	11.93	8.64	3.29
23	4.85		0.54	0.25	0.29	4.31	4.85	11.13	5.15	5.98
24	8.94		0.57	0.27	0.3	8.37	8.94	6.38	3.02	3.36
25	4.07		0.24	0.09	0.15	3.83	4.07	5.90	2.21	3.69
26	5.46		0.18	0.11	0.07	5.28	5.46	3.30	2.01	1.28
27	4.09	1.31	0.38	0.22	0.16	3.71	2.78	13.67	7.91	5.76
28	3.93	0.79	0.47	0.21	0.26	3.46	3.14	14.97	6.69	8.28
29	4.67	3.62	0.2	0.11	0.09	4.47	1.05	19.05	10.48	8.57
30	8.21	3.05	0.26	0.09	0.17	7.95	5.16	5.04	1.74	3.29
31	3.42	0.24	0.25	0.17	0.08	3.17	3.18	7.86	5.35	2.34
32	3.38	0.19	0.15	0.04	0.11	3.23	3.19	4.44	1.18	3.25
33	4.29	0.74	0.2	0.09	0.11	4.09	3.55	4.66	2.10	2.56
34	6.02	0.7	0.36	0.25	0.11	5.66	5.32	5.98	4.15	1.83
35	4.01	0.34	0.33	0.12	0.21	3.68	3.67	8.23	2.99	5.24
36	18.73	10.04	0.54	0.23	0.31	18.19	8.69	2.88	1.23	1.66
37	11.27	2.08	0.22	0.1	0.12	11.05	9.19	1.95	0.89	1.06
38	10.26	5.18	0.33	0.12	0.21	9.93	5.08	3.22	1.17	2.05
39	12.68	1.92	0.2	0.04	0.16	12.48	10.76	1.58	0.32	1.26
40	8.52	2.95	0.44	0.15	0.29	8.08	5.57	5.16	1.76	3.40
41	8.83	0.32	0.65	0.35	0.3	8.18	8.51	7.36	3.96	3.40
42	8.63	1.96	0.35	0.26	0.09	8.28	6.67	4.06	3.01	1.04
43	9.46	1.83	0.36	0.28	0.08	9.1	7.63	3.81	2.96	0.85
44	19.43	3.65	0.33	0.18	0.15	19.1	15.78	1.70	0.93	0.77
45	14.5	2.59	0.71	0.34	0.37	13.79	11.91	4.90	2.34	2.55
46	7.34	1.12	0.48	0.25	0.23	6.86	6.22	6.54	3.41	3.13
47	7.77	1.75	0.19	0.13	0.06	7.58	6.02	2.45	1.67	0.77
48	11.63	1.76	1.38	0.89	0.49	10.25	9.87	11.87	7.65	4.21
49	9.57	0.9	0.2	0.05	0.15	9.37	8.67	2.09	0.52	1.57
50	10.95	1.74	0.33	0.12	0.21	10.62	9.21	3.01	1.10	1.92
51	12.65	2.33	0.33	0.18	0.15	12.32	10.32	2.61	1.42	1.19
52	7.52	0.2	0.2	0.06	0.14	7.32	7.32	2.66	0.80	1.86

**Table A5: Line 5 - weekly parasitemia & gametocytemia by flow cytometry** (G=Gams, M=Males, F=Females, A=asexuals) before discounting for control readings in uninfected mouse (UI). Discounted data is shown as T' (total parasitemia), G' (gams), M' (males) & F' (females)

Weeks	Undiscounted Total (%)	Total in control	Discounted (as total %)				Discounted Total (%)	Discounted (as % of total)		
			G	M	F	A		G'	M'	F'
1	4.02		0.45	0.32	0.13	3.57	4.02	11.19	7.96	3.23
2	5.06		0.43	0.32	0.11	4.63	5.06	8.50	6.32	2.17
3	2.47		0.18	0.11	0.07	2.29	2.47	7.29	4.45	2.83
4	2.44		0.14	0.09	0.05	2.3	2.44	5.74	3.69	2.05
5	3.74		0.75	0.35	0.4	2.99	3.74	20.05	9.36	10.70
6	1.72		0.11	0.08	0.03	1.61	1.72	6.40	4.65	1.74
7	3.06		0.28	0.21	0.07	2.78	3.06	9.15	6.86	2.29
8	2.19		0.13	0.1	0.03	2.06	2.19	5.94	4.57	1.37
9	3.16		0.38	0.29	0.09	2.78	3.16	12.03	9.18	2.85
10	1.2		0.09	0.07	0.02	1.11	1.2	7.50	5.83	1.67
11	0.68		0.06	0.04	0.02	0.62	0.68	8.82	5.88	2.94
12	3.22		0.28	0.21	0.07	2.94	3.22	8.70	6.52	2.17
13	3.44		0.63	0.3	0.33	2.81	3.44	18.31	8.72	9.59
14	3.68		0.59	0.34	0.25	3.09	3.68	16.03	9.24	6.79
15	3.49		0.68	0.46	0.22	2.81	3.49	19.48	13.18	6.30
16	4.81		0.76	0.53	0.23	4.05	4.81	15.80	11.02	4.78
17	5		1.03	0.5	0.53	3.97	5	20.60	10.00	10.60
18	4.44		0.85	0.49	0.36	3.59	4.44	19.14	11.04	8.11
19	5.25		0.55	0.26	0.29	4.7	5.25	10.48	4.95	5.52
20	3.18		0.68	0.47	0.21	2.5	3.18	21.38	14.78	6.60
21	4.9		0.85	0.63	0.22	4.05	4.9	17.35	12.86	4.49
22	6.78		0.81	0.38	0.43	5.97	6.78	11.95	5.60	6.34
23	4.85		0.28	0.16	0.12	4.57	4.85	5.77	3.30	2.47
24	12.77		0.67	0.35	0.32	12.1	12.77	5.25	2.74	2.51
25	8.6		0.23	0.12	0.11	8.37	8.6	2.67	1.40	1.28
26	13.19	1.31	1.09	0.61	0.48	12.1	11.88	9.18	5.13	4.04
27	5.21	0.6	0.33	0.14	0.19	4.88	4.61	7.16	3.04	4.12
28	7.19	3.62	0.16	0.09	0.07	7.03	3.57	4.48	2.52	1.96
29	6.35	3.05	0.12	0.06	0.06	6.23	3.3	3.64	1.82	1.82
30	6.26	0.24	0.47	0.27	0.2	5.79	6.02	7.81	4.49	3.32
31	4.07	0.19	0.13	0.03	0.1	3.94	3.88	3.35	0.77	2.58
32	6.23	0.37	0.15	0.11	0.04	6.08	5.86	2.56	1.88	0.64
33	6.72	0.7	0.41	0.24	0.17	6.31	6.02	6.10	3.57	2.53
34	6.05	0.34	0.32	0.16	0.16	5.73	5.71	5.29	2.64	2.64
35	18.02	10.04	0.32	0.11	0.21	17.7	7.98	1.78	0.61	1.17
36	10.38	2.08	0.32	0.2	0.12	10.06	8.3	3.08	1.93	1.16
37	12.11	3.64	0.2	0.11	0.09	11.91	8.47	1.65	0.91	0.74
38	12.5	1.92	0.34	0.11	0.23	12.16	10.58	2.72	0.88	1.84
39	16.37	2.95	0.52	0.22	0.3	15.85	13.42	3.18	1.34	1.83
40	11.59	0.32	0.27	0.12	0.15	11.32	11.27	2.33	1.04	1.29
41	26.24	16.99	0.3	0.12	0.18	25.94	9.25	1.14	0.46	0.69
42	9.89	2.77	0.41	0.21	0.2	9.48	7.12	4.15	2.12	2.02
43	11.37	3.65	0.37	0.25	0.12	11	7.72	3.25	2.20	1.06
44	11.73	2.59	0.25	0.16	0.09	11.48	9.14	2.13	1.36	0.77
45	7.65	1.12	0.27	0.1	0.17	7.38	6.53	3.53	1.31	2.22
46	7.91	1.75	0.1	0.06	0.04	7.81	6.16	1.26	0.76	0.51
47	12.44	0.4	0.26	0.14	0.12	12.18	12.04	2.09	1.13	0.96
48	8.95	2.92	0.21	0.1	0.11	8.74	6.03	2.35	1.12	1.23
49	10.47	1.74	0.4	0.16	0.24	10.07	8.73	3.82	1.53	2.29
50	13.03	2.33	0.32	0.19	0.13	12.71	10.7	2.46	1.46	1.00
51	11.29	0.2	0.27	0.08	0.19	11.02	11.09	2.39	0.71	1.68



**Table A6: Line 6 - weekly parasitemia & gametocytemia by flow cytometry** (G=Gams, M=Males, F=Females, A=asexuals) before discounting for control readings in uninfected mouse (UI). Discounted data is shown as T' (total parasitemia), G' (gams), M' (males) & F' (females)

Weeks	Undiscounted Total (%)	Total in control (%)	Discounted (as total %)				Discounted Total (%)	Discounted (as % of total parasites)		
			G	M	F	A		G'	M'	F'
1	3.85		0.41	0.23	0.18	3.44	3.85	10.65	5.97	4.68
2	2.46		0.25	0.2	0.05	2.21	2.46	10.16	8.13	2.03
3	2.65		0.16	0.1	0.06	2.49	2.65	6.04	3.77	2.26
4	3.05		0.37	0.24	0.13	2.68	3.05	12.13	7.87	4.26
5	2.71		0.39	0.18	0.21	2.32	2.71	14.39	6.64	7.75
6	0.34		0.03	0.02	0.01	0.31	0.34	8.82	5.88	2.94
7	2.4		0.17	0.12	0.05	2.23	2.4	7.08	5.00	2.08
8	2.91		0.14	0.1	0.04	2.77	2.91	4.81	3.44	1.37
9	1.83		0.05	0.03	0.02	1.78	1.83	2.73	1.64	1.09
10	2.84		0.11	0.08	0.03	2.73	2.84	3.87	2.82	1.06
11	0.42		0.02	0.01	0.01	0.4	0.42	4.76	2.38	2.38
12	1.32		0.06	0.04	0.02	1.26	1.32	4.55	3.03	1.52
13	1.95		0.22	0.14	0.08	1.73	1.95	11.28	7.18	4.10
14	3.68		0.78	0.46	0.32	2.9	3.68	21.20	12.50	8.70
15	6.08		0.94	0.39	0.55	5.14	6.08	15.46	6.41	9.05
16	4.31		0.67	0.24	0.43	3.64	4.31	15.55	5.57	9.98
17	5.19		1.13	0.54	0.59	4.06	5.19	21.77	10.40	11.37
18	3.79		0.86	0.53	0.33	2.93	3.79	22.69	13.98	8.71
19	4.48		0.37	0.27	0.1	4.11	4.48	8.26	6.03	2.23
20	4.78		0.51	0.34	0.17	4.27	4.78	10.67	7.11	3.56
21	2.96		0.19	0.11	0.08	2.77	2.96	6.42	3.72	2.70
22	3.7		0.6	0.42	0.18	3.1	3.7	16.22	11.35	4.86
23	3.92		0.62	0.36	0.26	3.3	3.92	15.82	9.18	6.63
24	3.37		1.23	0.6	0.63	2.14	3.37	36.50	17.80	18.69
25	5.71		0.55	0.36	0.19	5.16	5.71	9.63	6.30	3.33
26	8.19		0.53	0.27	0.26	7.66	8.19	6.47	3.30	3.17
27	10.55		0.71	0.29	0.42	9.84	10.55	6.73	2.75	3.98
28	11.84		0.2	0.1	0.1	11.64	11.84	1.69	0.84	0.84
29	5.47		0.33	0.19	0.14	5.14	5.47	6.03	3.47	2.56
30	5.17	1.77	0.26	0.18	0.08	4.91	3.4	7.65	5.29	2.35
31	7.27	0.79	0.34	0.13	0.21	6.93	6.48	5.25	2.01	3.24
32	5.46	3.62	0.26	0.09	0.17	5.2	1.84	14.13	4.89	9.24
33	7.14	3.05	0.28	0.1	0.18	6.86	4.09	6.85	2.44	4.40
34	5.89	0.2	0.46	0.15	0.31	5.43	5.69	8.08	2.64	5.45
35	4.34	0.19	0.13	0.04	0.09	4.21	4.15	3.13	0.96	2.17
36	4.76	0.27	0.31	0.09	0.22	4.45	4.49	6.90	2.00	4.62
37	6.66	0.7	0.22	0.14	0.08	6.44	5.96	3.30	2.10	1.20
38	6.99	0.34	0.39	0.11	0.28	6.6	6.65	5.58	1.57	4.01
39	14.9	6.74	0.31	0.23	0.08	14.59	8.16	2.08	1.54	0.54
40	8.85	2.08	0.22	0.13	0.09	8.63	6.77	2.49	1.47	1.02
41	9.12	3.64	0.12	0.08	0.04	9	5.48	1.32	0.88	0.44
42	10.76	1.92	0.22	0.07	0.15	10.54	8.84	2.04	0.65	1.39
43	15.56	2.95	0.18	0.07	0.11	15.38	12.61	1.16	0.45	0.71
44	11.71	0.32	0.44	0.21	0.23	11.27	11.39	3.76	1.79	1.96
45	26.63	16.99	0.48	0.18	0.3	26.15	9.64	1.80	0.68	1.13
46	9.21	1.83	0.26	0.2	0.06	8.95	7.38	2.82	2.17	0.65
47	14.95	3.65	0.28	0.15	0.13	14.67	11.3	1.87	1.00	0.87
48	12.37	2.59	0.32	0.2	0.12	12.05	9.78	2.59	1.62	0.97
49	7.78	1.12	0.33	0.13	0.2	7.45	6.66	4.24	1.67	2.57
50	17.8	0.63	0.25	0.09	0.16	17.55	17.17	1.40	0.51	0.90
51	8.33	0.4	0.5	0.26	0.24	7.83	7.93	6.00	3.12	2.88
52	8.18	0.31	0.25	0.09	0.16	7.93	7.87	3.06	1.10	1.96
53	12.17	1.74	0.44	0.17	0.27	11.73	10.43	3.62	1.40	2.22

**Table A7: Line 7 - weekly parasitemia & gametocytemia by flow cytometry** (G=Gams, M=Males, F=Females, A=asexuals) before discounting for control readings in uninfected mouse (UI). Discounted data is shown as T' (total parasitemia), G' (gams), M' (males) & F' (females)

Weeks	Undiscounted Total (%)	Total in control (%)	Discounted (as total %)				Discounted Total (%)	Discounted (as % of total parasites)		
			G	M	F	A		G'	M'	F'
1	4.63		0.43	0.23	0.2	4.2	4.63	9.29	4.97	4.32
2	1.92		0.16	0.14	0.02	1.76	1.92	8.33	7.29	1.04
3	1.01		0.08	0.05	0.03	0.93	1.01	7.92	4.95	2.97
4	7.51		0.88	0.4	0.48	6.63	7.51	11.72	5.33	6.39
5	4.2		0.19	0.14	0.05	4.01	4.2	4.52	3.33	1.19
6	3.81		0.3	0.2	0.1	3.51	3.81	7.87	5.25	2.62
7	2.49		0.11	0.08	0.03	2.38	2.49	4.42	3.21	1.20
8	3.96		0.5	0.23	0.27	3.46	3.96	12.63	5.81	6.82
9	0.13		0	0	0	0.13	0.13	0.00	0.00	0.00
10	4.71		0.35	0.19	0.16	4.36	4.71	7.43	4.03	3.40
11	5.07		0.33	0.24	0.09	4.74	5.07	6.51	4.73	1.78
12	6.01		0.46	0.31	0.15	5.55	6.01	7.65	5.16	2.50
13	5.79		1.33	0.68	0.65	4.46	5.79	22.97	11.74	11.23
14	9.17		1.05	0.47	0.58	8.12	9.17	11.45	5.13	6.32
15	6.86		0.66	0.28	0.38	6.2	6.86	9.62	4.08	5.54
16	3.64		0.24	0.17	0.07	3.4	3.64	6.59	4.67	1.92
17	7.72		0.73	0.35	0.38	6.99	7.72	9.46	4.53	4.92
18	6.94		0.95	0.48	0.47	5.99	6.94	13.69	6.92	6.77
19	3.27		0.25	0.17	0.08	3.02	3.27	7.65	5.20	2.45
20	4.4		0.37	0.27	0.1	4.03	4.4	8.41	6.14	2.27
21	6.74		0.65	0.44	0.21	6.09	6.74	9.64	6.53	3.12
22	5.85		0.78	0.31	0.47	5.07	5.85	13.33	5.30	8.03
23	5.16		0.58	0.4	0.18	4.58	5.16	11.24	7.75	3.49
24	10.26		0.43	0.21	0.22	9.83	10.26	4.19	2.05	2.14
25	6.41		0.09	0.02	0.07	6.32	6.41	1.40	0.31	1.09
26	13.92	8.72	0.03	0.01	0.02	13.89	5.2	0.58	0.19	0.38
27	11.04	1.7	0	0	0	11.04	9.34	0.00	0.00	0.00
28	6.64	1.77	0	0	0	6.64	4.87	0.00	0.00	0.00
29	6.87	0.79	0	0	0	6.87	6.08	0.00	0.00	0.00
30	11.57	3.62	0	0	0	11.57	7.95	0.00	0.00	0.00
31	14.78	3.05	0	0	0	14.78	11.73	0.00	0.00	0.00
32	4.46	0.24	0	0	0	4.46	4.22	0.00	0.00	0.00
33	9.56	0.23	0	0	0	9.56	9.33	0.00	0.00	0.00
34	18.74	0.74	0	0	0	18.74	18	0.00	0.00	0.00

**Table A8: Line 8 - weekly parasitemia & gametocytemia by flow cytometry** (G=Gams, M=Males, F=Females, A=asexuals) before discounting for control readings in uninfected mouse (UI). Discounted data is shown as T' (total parasitemia), G' (gams), M' (males) & F' (females)

Weeks	Undiscounted Total (%)	Total in control (%)	Discounted (as total %)				Discounted Total (%)	Discounted (as % of total parasites)		
			G	M	F	A		G'	M'	F'
1	5.55		0.43	0.28	0.15	5.12	5.55	7.75	5.05	2.70
2	3.02		0.2	0.15	0.05	2.82	3.02	6.62	4.97	1.66
3	2.75		0.19	0.12	0.07	2.56	2.75	6.91	4.36	2.55
4	3.66		0.6	0.28	0.32	3.06	3.66	16.39	7.65	8.74
5	2.88		0.75	0.32	0.43	2.13	2.88	26.04	11.11	14.93
6	2.39		0.13	0.1	0.03	2.26	2.39	5.44	4.18	1.26
7	3.81		0.34	0.2	0.14	3.47	3.81	8.92	5.25	3.67
8	3.54		0.18	0.13	0.05	3.36	3.54	5.08	3.67	1.41
9	1.96		0.13	0.09	0.04	1.83	1.96	6.63	4.59	2.04
10	1.86		0.13	0.1	0.03	1.73	1.86	6.99	5.38	1.61
11	0.8		0.03	0.02	0.01	0.77	0.8	3.75	2.50	1.25
12	2.9		0.2	0.13	0.07	2.7	2.9	6.90	4.48	2.41
13	5.39		0.55	0.22	0.33	4.84	5.39	10.20	4.08	6.12
14	5.34		0.58	0.37	0.21	4.76	5.34	10.86	6.93	3.93
15	4.55		0.8	0.33	0.47	3.75	4.55	17.58	7.25	10.33
16	5.02		0.82	0.53	0.29	4.2	5.02	16.33	10.56	5.78
17	3.81		0.3	0.16	0.14	3.51	3.81	7.87	4.20	3.67
18	8.72		0.52	0.24	0.28	8.2	8.72	5.96	2.75	3.21
19	3.67		0.27	0.2	0.07	3.4	3.67	7.36	5.45	1.91
20	6.09		0.68	0.4	0.28	5.41	6.09	11.17	6.57	4.60
21	14.53		1.04	0.5	0.54	13.49	14.53	7.16	3.44	3.72
22	15.01		0.57	0.26	0.31	14.44	15.01	3.80	1.73	2.07
23	7.72		0.46	0.23	0.23	7.26	7.72	5.96	2.98	2.98
24	9.57		0.57	0.26	0.31	9	9.57	5.96	2.72	3.24
25	8.59		0.33	0.13	0.2	8.26	8.59	3.84	1.51	2.33
26	13.11		0.44	0.25	0.19	12.67	13.11	3.36	1.91	1.45
27	8.54		0.36	0.21	0.15	8.18	8.54	4.22	2.46	1.76
28	6.23	1.77	0.34	0.15	0.19	5.89	4.46	7.62	3.36	4.26
29	6.71	0.79	0.43	0.16	0.27	6.28	5.92	7.26	2.70	4.56
30	5.9	3.62	0.29	0.13	0.16	5.61	2.28	12.72	5.70	7.02
31	5.4	3.05	0.11	0.05	0.06	5.29	2.35	4.68	2.13	2.55
32	4.96	0.24	0.25	0.19	0.06	4.71	4.72	5.30	4.03	1.27
33	3.38	0.19	0.08	0.06	0.02	3.3	3.19	2.51	1.88	0.63
34	4.74	0.27	0.37	0.14	0.23	4.37	4.47	8.28	3.13	5.15
35	7.34	0.7	0.36	0.27	0.09	6.98	6.64	5.42	4.07	1.36
36	4.37	0.34	0.18	0.11	0.07	4.19	4.03	4.47	2.73	1.74
37	22	10.04	0.29	0.26	0.03	21.71	11.96	2.42	2.17	0.25
38	8.74	2.69	0.13	0.05	0.08	8.61	6.05	1.49	0.57	0.92
39	10.43	5.18	0.32	0.11	0.21	10.11	5.25	3.07	1.05	2.01
40	9.94	1.92	0.26	0.07	0.19	9.68	8.02	2.62	0.70	1.91
41	9.45	2.95	0.27	0.11	0.16	9.18	6.5	2.86	1.16	1.69
42	8.2	0.32	0.19	0.1	0.09	8.01	7.88	2.32	1.22	1.10
43	31.94	16.99	0.64	0.34	0.3	31.3	14.95	2.00	1.06	0.94
44	8.64	1.83	0.18	0.11	0.07	8.46	6.81	2.08	1.27	0.81
45	11.48	3.65	0.18	0.08	0.1	11.3	7.83	1.57	0.70	0.87
46	10.96	2.59	0.17	0.09	0.08	10.79	8.37	1.55	0.82	0.73
47	4.28	0.43	0.05	0.01	0.04	4.23	3.85	1.17	0.23	0.93
48	6.83	0.85	0.05	0.03	0.02	6.78	5.98	0.73	0.44	0.29
49	5.12	0.71	0	0	0	5.12	4.41	0.00	0.00	0.00
50	9.57	0.3	0	0	0	9.57	9.27	0.00	0.00	0.00
51	13.73	1.74	0	0	0	13.73	11.99	0.00	0.00	0.00
52	10.75	1.77	0	0	0	10.75	8.98	0.00	0.00	0.00
53	7.24	0.17	0	0	0	7.24	7.07	0.00	0.00	0.00

**Table A9: Line 9 - weekly parasitemia & gametocytemia by flow cytometry** (G=Gams, M=Males, F=Females, A=asexuals) before discounting for control readings in uninfected mouse (UI). Discounted data is shown as T' (total parasitemia), G' (gams), M' (males) & F' (females)

Weeks	Undiscounted Total (%)	Total in control (%)	Discounted (as total %)				Discounted Total (%)	Discounted (as % of total parasites)		
			G	M	F	A		G'	M'	F'
1	4.49		0.43	0.24	0.19	4.06	4.49	9.58	5.35	4.23
2	0.9		0.06	0.05	0.01	0.84	0.9	6.67	5.56	1.11
3	0.97		0.05	0.03	0.02	0.92	0.97	5.15	3.09	2.06
4	2.14		0.12	0.08	0.04	2.02	2.14	5.61	3.74	1.87
5	3.45		0.74	0.36	0.38	2.71	3.45	21.45	10.43	11.01
6	3.86		0.44	0.31	0.13	3.42	3.86	11.40	8.03	3.37
7	3.82		0.47	0.32	0.15	3.35	3.82	12.30	8.38	3.93
8	3.58		0.22	0.15	0.07	3.36	3.58	6.15	4.19	1.96
9	0.91		0.03	0.02	0.01	0.88	0.91	3.30	2.20	1.10
10	3.32		0.52	0.3	0.22	2.8	3.32	15.66	9.04	6.63
11	2.55		0.45	0.17	0.28	2.1	2.55	17.65	6.67	10.98
12	3.06		0.17	0.11	0.06	2.89	3.06	5.56	3.59	1.96
13	4.42		0.23	0.13	0.1	4.19	4.42	5.20	2.94	2.26
14	3.89		0.22	0.1	0.12	3.67	3.89	5.66	2.57	3.08
15	2.8		0.1	0.07	0.03	2.7	2.8	3.57	2.50	1.07
16	2.34		0.02	0.01	0.01	2.32	2.34	0.85	0.43	0.43
17	3.25		0.03	0.02	0.01	3.22	3.25	0.92	0.62	0.31
18	3.78	0.67	0	0	0	3.78	3.11	0.00	0.00	0.00
19	3.33	0.09	0	0	0	3.33	3.24	0.00	0.00	0.00
20	5.05	0.11	0	0	0	5.05	4.94	0.00	0.00	0.00
21	5.11	0.22	0	0	0	5.11	4.89	0.00	0.00	0.00

**Table A10: Line 10 - weekly parasitemia & gametocytemia by flow cytometry** (G=Gams, M=Males, F=Females, A=asexuals) before discounting for control readings in uninfected mouse (UI). Discounted data is shown as T' (total parasitemia), G' (gams), M' (males) & F' (females)

Weeks	Undicounted Total (%)	Total in control (%)	Discounted (as total %)				Discounted Total (%)	Discounted (as % of total parasites)		
			G	M	F	A		G'	M'	F'
1	5.83		0.41	0.17	0.24	5.42	5.83	7.03	2.92	4.12
2	2.7		0.18	0.14	0.04	2.52	2.7	6.67	5.19	1.48
3	2.81		0.15	0.11	0.04	2.66	2.81	5.34	3.91	1.42
4	6.62		0.77	0.48	0.29	5.85	6.62	11.63	7.25	4.38
5	7.44		0.8	0.38	0.42	6.64	7.44	10.75	5.11	5.65
6	5.4		0.43	0.26	0.17	4.97	5.4	7.96	4.81	3.15
7	2.09		0.11	0.07	0.04	1.98	2.09	5.26	3.35	1.91
8	3.38		0.22	0.15	0.07	3.16	3.38	6.51	4.44	2.07
9	1.06		0.04	0.03	0.01	1.02	1.06	3.77	2.83	0.94
10	3.77		0.2	0.13	0.07	3.57	3.77	5.31	3.45	1.86
11	0.21		0.01	0.01	0	0.2	0.21	4.76	4.76	0.00
12	4.73		0.63	0.27	0.36	4.1	4.73	13.32	5.71	7.61
13	3.12		0.33	0.22	0.11	2.79	3.12	10.58	7.05	3.53
14	7.03		0.29	0.15	0.14	6.74	7.03	4.13	2.13	1.99
15	7.46		0.15	0.08	0.07	7.31	7.46	2.01	1.07	0.94
16	4.69		0.53	0.3	0.23	4.16	4.69	11.30	6.40	4.90
17	5.31		0.46	0.22	0.24	4.85	5.31	8.66	4.14	4.52
18	4.31		0.3	0.22	0.08	4.01	4.31	6.96	5.10	1.86
19	5.87		0.76	0.44	0.32	5.11	5.87	12.95	7.50	5.45
20	7.08		0.57	0.26	0.31	6.51	7.08	8.05	3.67	4.38
21	5.47		0.52	0.4	0.12	4.95	5.47	9.51	7.31	2.19
22	4.47		0.74	0.55	0.19	3.73	4.47	16.55	12.30	4.25
23	6.25		0.51	0.21	0.3	5.74	6.25	8.16	3.36	4.80
24	8.77		1.61	0.9	0.71	7.16	8.77	18.36	10.26	8.10
25	6.18		0.61	0.4	0.21	5.57	6.18	9.87	6.47	3.40
26	10.47		0.68	0.33	0.35	9.79	10.47	6.49	3.15	3.34
27	8.57		0.51	0.19	0.32	8.06	8.57	5.95	2.22	3.73
28	12.69		0.45	0.28	0.17	12.24	12.69	3.55	2.21	1.34
29	8.84		0.26	0.11	0.15	8.58	8.84	2.94	1.24	1.70
30	6.16	1.77	0.21	0.09	0.12	5.95	4.39	4.78	2.05	2.73
31	7.28	0.89	0.15	0.08	0.07	7.13	6.39	2.35	1.25	1.10
32	7.15	3.62	0.12	0.07	0.05	7.03	3.53	3.40	1.98	1.42
33	14.56	3.05	0.07	0.04	0.03	14.49	11.51	0.61	0.35	0.26
34	4.57	1.75	0.12	0.07	0.05	4.45	2.82	4.26	2.48	1.77
35	4.92	0.23	0.04	0.03	0.01	4.88	4.69	0.85	0.64	0.21
36	5.33	0.27	0.13	0.07	0.06	5.2	5.06	2.57	1.38	1.19
37	10.56	0.7	0.59	0.5	0.09	9.97	9.86	5.98	5.07	0.91
38	10.65	0.34	0.66	0.51	0.15	9.99	10.31	6.20	4.79	1.41
39	15.79	10.04	0.23	0.08	0.15	15.56	5.75	1.46	0.51	0.95
40	12.91	2.69	0.31	0.15	0.16	12.6	10.22	2.40	1.16	1.24
41	9.16	5.18	0.25	0.06	0.19	8.91	3.98	2.73	0.66	2.07
42	8.74	1.92	0.27	0.11	0.16	8.47	6.82	3.09	1.26	1.83
43	8.46	2.95	0.28	0.07	0.21	8.18	5.51	3.31	0.83	2.48
44	9.55	0.32	0.19	0.07	0.12	9.36	9.23	1.99	0.73	1.26
45	7.35	1.96	0.16	0.1	0.06	7.19	5.39	2.18	1.36	0.82
46	7.53	2.77	0.21	0.08	0.13	7.32	4.76	2.79	1.06	1.73
47	10.79	3.65	0.17	0.08	0.09	10.62	7.14	1.58	0.74	0.83
48	10.17	2.59	0.26	0.14	0.12	9.91	7.58	2.56	1.38	1.18
49	6.23	1.12	0.27	0.09	0.18	5.96	5.11	4.33	1.44	2.89
50	6.26	1.75	0.08	0.02	0.06	6.18	4.51	1.28	0.32	0.96
51	9.56	1.76	0.75	0.52	0.23	8.81	7.8	7.85	5.44	2.41
52	9.57	2.92	0.16	0.08	0.08	9.41	6.65	1.67	0.84	0.84
53	7.69	1.11	0.24	0.11	0.13	7.45	6.58	3.12	1.43	1.69

**Table A11 – Details of mutated gene products temporarily considered under reduced ranking priority and hence not included in the analysis**

REF STR	LINE	TYPE	CHR	GENE ID	PRODUCT
COMMON IN ALL GNPs					
ANKA	233	SNP	2	PBANKA_020060	Pb-fam-1 protein"
ANKA	M7	SNP	2	PBANKA_020060	Pb-fam-1 protein"
ANKA	M9	SNP	2	PBANKA_020060	Pb-fam-1 protein"
COMMON IN 233, M7 AND M9					
ANKA	M7	INDEL	3	PBANKA_030330	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	INDEL	3	PBANKA_030330	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	233	SNP - 1	5	PBANKA_050025	tryptophan-rich antigen, putative, pseudogene"
ANKA	M7	SNP - 3	5	PBANKA_050025	tryptophan-rich antigen, putative, pseudogene"
ANKA	M9	SNP - 2	5	PBANKA_050025	tryptophan-rich antigen, putative, pseudogene"
COMMON IN M7 AND M9					
820	M7	INDEL	11	PBANKA_114510	conserved <i>Plasmodium</i> protein, unknown function"
820	M9	INDEL	11	PBANKA_114510	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M7	INDEL	2	PBANKA_021270	zinc finger protein, putative"
ANKA	M9	INDEL	2	PBANKA_021270	zinc finger protein, putative"
ANKA	M7	SNP - 4	2	PBANKA_020090	conserved rodent malaria protein, unknown
ANKA	M9	SNP	2	PBANKA_020090	conserved rodent malaria protein, unknown
ANKA	M7	INDEL	12	PBANKA_122610	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	INDEL - 2	12	PBANKA_122610	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M7	INDEL	13	PBANKA_131030	conserved <i>Plasmodium</i> protein, unknown function,
ANKA	M9	INDEL	13	PBANKA_131030	conserved <i>Plasmodium</i> protein, unknown function,
ANKA	M7	SNP	14	PBANKA_146500	PYST-C2 homologue, putative"
ANKA	M9	SNP	14	PBANKA_146500	PYST-C2 homologue, putative"
ANKA	234	SNP	14	PBANKA_146500	PYST-C2 homologue, putative"
ANKA	M7	INDEL	1	PBANKA_011180	adapter-related protein, putative"
ANKA	M9	INDEL	1	PBANKA_011180	adapter-related protein, putative"
ANKA	M7	INDEL	2	PBANKA_020700	calcium-transporting ATPase, putative"
ANKA	M9	INDEL	2	PBANKA_020700	calcium-transporting ATPase, putative"
ANKA	M7	INDEL	2	PBANKA_020710	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	INDEL	2	PBANKA_020710	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M7	INDEL	2	PBANKA_021270	zinc finger protein, putative"
ANKA	M9	INDEL	2	PBANKA_021270	zinc finger protein, putative"
ANKA	M7	INDEL	3	PBANKA_030330	conserved <i>Plasmodium</i> protein, unknown function"

REF STR	LINE	TYPE	CHR	GENE ID	PRODUCT
ANKA	M9	INDEL	3	PBANKA_030330	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M7	INDEL	3	PBANKA_030400	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	INDEL	3	PBANKA_030400	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M7	INDEL - 1	12	PBANKA_122610	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	INDEL - 2	12	PBANKA_122610	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M7	INDEL	13	PBANKA_131030	conserved <i>Plasmodium</i> protein, unknown function,
ANKA	M9	INDEL	13	PBANKA_131030	conserved <i>Plasmodium</i> protein, unknown function,
COMMON IN M7 AND M9					
ANKA	233	SNP - 2	5	PBANKA_050020	BIR protein"
ANKA	M7	SNP	5	PBANKA_050020	BIR protein"
ANKA	233	SNP - 3	B	PBANKA_000440	BIR protein, fragment"
ANKA	M7	SNP	B	PBANKA_000440	BIR protein, fragment"
COMMON IN 233 AND M9					
820	233	SNP - 3	5	PBANKA_050070	Pb-fam-1 protein
820	M9	SNP - 2	5	PBANKA_050070	Pb-fam-1 protein
ANKA	233	SNP - 2	B	PBANKA_000550	<i>Plasmodium</i> exported protein, unknown function"
ANKA	M9	SNP - 2	B	PBANKA_000550	<i>Plasmodium</i> exported protein, unknown function"
ANKA	233	INDEL	3	PBANKA_030240	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	INDEL	3	PBANKA_030240	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	234	INDEL	3	PBANKA_030240	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	233	INDEL	3	PBANKA_030250	hexose transporter"
ANKA	M9	INDEL	3	PBANKA_030250	hexose transporter"
ANKA	234	INDEL	3	PBANKA_030250	hexose transporter"
ANKA	233	SNP - 1 INDEL - 1	7	PBANKA_072260	<i>Plasmodium</i> exported protein, unknown function"
ANKA	M9	SNP - 2	7	PBANKA_072260	<i>Plasmodium</i> exported protein, unknown function"
ANKA	233	SNP	9	PBANKA_090090	reticulocyte binding protein, putative,
ANKA	M9	SNP	9	PBANKA_090090	reticulocyte binding protein, putative,
ANKA	820	INDEL	9	PBANKA_090090	reticulocyte binding protein, putative,
ANKA	233	SNP - 2	10	PBANKA_100060	erythrocyte membrane antigen 1"
ANKA	M9	SNP - 2	10	PBANKA_100060	erythrocyte membrane antigen 1"
ANKA	233	INDEL	11	PBANKA_113040	proteasome subunit, putative"
ANKA	M9	INDEL	11	PBANKA_113040	proteasome subunit, putative"
ANKA	233	SNP	14	PBANKA_140030	BIR protein"
ANKA	M9	SNP	14	PBANKA_140030	BIR protein"
ANKA	233	INDEL	14	PBANKA_144600	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	INDEL	14	PBANKA_144600	conserved <i>Plasmodium</i> protein, unknown function"

233 ALONE					
REF STR	LINE	TYPE	CHR	GENE ID	PRODUCT
820	233	SNP	1	PBANKA_011220	myosin-like protein, putative"
820	233	SNP	3	PBANKA_031340	conserved <i>Plasmodium</i> protein, unknown function"
820	233	SNP	4	PBANKA_041130	conserved <i>Plasmodium</i> protein, unknown function"
820	233	SNP	5	PBANKA_051540	conserved <i>Plasmodium</i> protein, unknown function"
820	233	SNP	6	PBANKA_061220	conserved <i>Plasmodium</i> protein, unknown function"
820	233	SNP - 2	9	PBANKA_090230	conserved <i>Plasmodium</i> protein, unknown function,
820	233	SNP - 2	9	PBANKA_090240	conserved <i>Plasmodium</i> protein, unknown function,
820	233	INDEL	11	PBANKA_112960	conserved <i>Plasmodium</i> protein, unknown function"
820	233	INDEL	12	PBANKA_121660	conserved <i>Plasmodium</i> protein, unknown function"
820	233	INDEL	12	PBANKA_121680	conserved <i>Plasmodium</i> protein, unknown function"
820	233	SNP	12	PBANKA_124630	BIR protein"
820	233	SNP	14	PBANKA_143540	conserved <i>Plasmodium</i> protein, unknown function"
820	233	SNP	14	PBANKA_144010	conserved <i>Plasmodium</i> protein, unknown function"
820	233	INDEL -2	14	PBANKA_145320	conserved <i>Plasmodium</i> protein, unknown function"
820	233	SNP	14	PBANKA_146520	BIR protein"
820	233	SNP - 2	B	PBANKA_000200	BIR protein"
820	233	SNP - 4	B	PBANKA_000320	BIR protein"
ANKA	233	SNP - 5	2	PBANKA_021570	BIR protein"
ANKA	233	SNP	2	PBANKA_021540	<i>Plasmodium</i> exported protein, unknown function"
ANKA	233	SNP	4	PBANKA_040060	BIR protein"
ANKA	233	SNP	6	PBANKA_062390	BIR protein"
ANKA	233	SNP	6	PBANKA_062300	<i>Plasmodium</i> exported protein, unknown function"
ANKA	233	SNP	7	PBANKA_070010	BIR protein"
ANKA	233	SNP	7	PBANKA_070050	Pb-fam-1 protein"
ANKA	233	SNP	9	PBANKA_090010	conserved rodent malaria protein, unknown
ANKA	233	INDEL	10	PBANKA_102650	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	233	SNP	10	PBANKA_104050	BIR protein, pseudogene"
ANKA	233	SNP	14	PBANKA_143540	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	233	SNP	14	PBANKA_145310	sphingomyelin phosphodiesterase, putative"
ANKA	233	INDEL	14	PBANKA_145320	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	233	SNP - 3	B	PBANKA_000100	BIR protein"
ANKA	233	SNP	B	PBANKA_000300	BIR protein, fragment"
ANKA	233	SNP - 2	B	PBANKA_000340	BIR protein"



REF STR	LINE	TYPE	CHR	GENE ID	PRODUCT
ANKA	233	SNP - 2		PBANKA_000300	BIR protein, fragment"
ANKA	233	SNP - 3		PBANKA_000310	BIR protein"
ANKA	233	SNP		PBANKA_000360	BIR protein"
ANKA	233	SNP - 5		PBANKA_000400	<i>Plasmodium</i> exported protein, unknown function"
ANKA	233	SNP - 3		PBANKA_000440	BIR protein, fragment"
ANKA	233	SNP - 7	2	PBANKA_021580	<i>Plasmodium</i> exported protein, unknown function"
ANKA	233	SNP - 2	2	PBANKA_021590	BIR protein, pseudogene"
ANKA	233	INDEL	3	PBANKA_030110	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	233	INDEL	3	PBANKA_030120	ERCC1 nucleotide excision repair protein,
ANKA	233	INDEL - 2	3	PBANKA_031570	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	233	SNP	4	PBANKA_040030	<i>Plasmodium</i> exported protein, unknown function"
ANKA	233	SNP	6	PBANKA_062300	<i>Plasmodium</i> exported protein, unknown function"
ANKA	233	SNP - 2	6	PBANKA_062390	BIR protein"
ANKA	233	SNP - 2	7	PBANKA_070010	BIR protein"
ANKA	233	SNP	7	PBANKA_072270	BIR protein, pseudogene"
ANKA	233	INDEL	8	PBANKA_083360	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	233	SNP	9	PBANKA_090010	conserved rodent malaria protein, unknown
ANKA	233	INDEL	10	PBANKA_102650	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	233	INDEL	12	PBANKA_121680	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	233	SNP	12	PBANKA_124620	<i>Plasmodium</i> exported protein, unknown function"
ANKA	233	SNP	13	PBANKA_130050	tubulin, putative"
ANKA	233	INDEL	13	PBANKA_132220	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	233	INDEL	13	PBANKA_132230	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	233	INDEL	13	PBANKA_133720	SFT2-like protein, putative"
ANKA	233	INDEL	13	PBANKA_136510	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	233	INDEL	14	PBANKA_141440	exportin-T, putative"
ANKA	233	INDEL	14	PBANKA_141450	protein kinase, putative"
ANKA	233	INDEL	14	PBANKA_142810	asparagine-rich antigen, putative"
ANKA	233	SNP	14	PBANKA_143540	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	233	INDEL	14	PBANKA_145060	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	233	SNP	14	PBANKA_145310	sphingomyelin phosphodiesterase, putative"
ANKA	233	INDEL	14	PBANKA_145320	conserved <i>Plasmodium</i> protein, unknown function"
M7 ALONE					
820	M7	INDEL	2	PBANKA_021270	zinc finger protein, putative"

REF STR	LINE	TYPE	CHR	GENE ID	PRODUCT
820	M7	INDEL	7	PBANKA_070510	conserved <i>Plasmodium</i> protein, unknown function"
820	M7	INDEL	8	PBANKA_082260	conserved <i>Plasmodium</i> protein, unknown function"
820	M7	INDEL	9	PBANKA_091000	conserved <i>Plasmodium</i> protein, unknown function"
820	M7	INDEL	10	PBANKA_100690	conserved <i>Plasmodium</i> protein, unknown function"
820	M7	INDEL	10	PBANKA_101970	conserved <i>Plasmodium</i> protein, unknown function"
820	M7	INDEL	10	PBANKA_102330	conserved <i>Plasmodium</i> protein, unknown function"
820	M7	INDEL	12	PBANKA_124580	conserved <i>Plasmodium</i> protein, unknown function"
820	M7	INDEL	13	PBANKA_131420	conserved <i>Plasmodium</i> protein, unknown function"
820	M7	INDEL	14	PBANKA_141260	telomerase reverse transcriptase, putative,
820	M7	INDEL	14	PBANKA_141270	telomerase reverse transcriptase, putative,
820	M7	INDEL	14	PBANKA_145190	conserved <i>Plasmodium</i> protein, unknown function"
820	M7	INDEL	11	PBANKA_112690	protein kinase, putative"
ANKA	M7	INDEL	3	PBANKA_031120	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M7	INDEL	5	PBANKA_051310	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M7	INDEL	7	PBANKA_070310	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M7	INDEL	8	PBANKA_080270	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M7	INDEL	8	PBANKA_082300	ubiquitin-like protein, putative"
ANKA	M7	INDEL	10	PBANKA_100690	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M7	INDEL	10	PBANKA_102330	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M7	SNP	12	PBANKA_120060	<i>Plasmodium</i> exported protein, unknown function"
ANKA	M7	INDEL	12	PBANKA_122590	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M7	INDEL	12	PBANKA_123350	RNA-binding protein, putative"
ANKA	M7	INDEL	12	PBANKA_124580	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M7	INDEL	14	PBANKA_141810	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M7	SNP	14	PBANKA_145190	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M7	SNP -4	B	PBANKA_000260	BIR protein"
ANKA	M7	INDEL	3	PBANKA_031120	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M7	INDEL	4	PBANKA_040290	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M7	INDEL	4	PBANKA_040300	60S ribosomal protein L44, putative"
ANKA	M7	INDEL	5	PBANKA_051310	conserved <i>Plasmodium</i> protein, unknown function"

REF STR	LINE	TYPE	CHR	GENE ID	PRODUCT
ANKA	M7	INDEL	7	PBANKA_070310	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M7	INDEL	8	PBANKA_080270	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M7	INDEL	8	PBANKA_082300	ubiquitin-like protein, putative"
ANKA	M7	INDEL	9	PBANKA_092830	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M7	INDEL	9	PBANKA_093020	DNA mismatch repair protein, putative"
ANKA	M7	INDEL	10	PBANKA_100690	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M7	INDEL	10	PBANKA_101730	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M7	INDEL	10	PBANKA_101740	selenoprotein, putative"
ANKA	M7	INDEL	10	PBANKA_102660	GTP-binding protein, putative"
ANKA	M7	INDEL	10	PBANKA_102670	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M7	INDEL	11	PBANKA_110030	BIR protein, pseudogene"
ANKA	M7	INDEL	11	PBANKA_110590	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M7	INDEL	11	PBANKA_110600	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M7	INDEL	11	PBANKA_113760	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M7	INDEL	11	PBANKA_114230	rhopty protein, putative"
ANKA	M7	INDEL	12	PBANKA_122590	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M7	INDEL	12	PBANKA_123450	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M7	INDEL	12	PBANKA_123460	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M7	INDEL	12	PBANKA_123780	multidrug resistance protein, putative"
ANKA	M7	INDEL	12	PBANKA_123790	mitochondrial processing peptidase alpha
ANKA	M7	INDEL	12	PBANKA_124360	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M7	INDEL	13	PBANKA_130310	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M7	INDEL	13	PBANKA_130320	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M7	INDEL	13	PBANKA_133830	60S ribosomal protein L6, putative"
ANKA	M7	INDEL	13	PBANKA_134880	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M7	INDEL	14	PBANKA_141810	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M7	INDEL	14	PBANKA_143830	phospholipid-transporting ATPase, putative"
ANKA	M7	INDEL	14	PBANKA_143840	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M7	INDEL	14	PBANKA_144650	asparagine-rich protein, putative"
ANKA	M7	INDEL	14	PBANKA_145180	conserved <i>Plasmodium</i> protein, unknown function"
<b>M9 ALONE</b>					
820	M9	SNP - 2	B	PBANKA_001030	BIR protein"

REF STR	LINE	TYPE	CHR	GENE ID	PRODUCT
820	M9	SNP	7	PBANKA_072270	BIR protein, pseudogene"
820	M9	INDEL	10	PBANKA_101430	conserved <i>Plasmodium</i> protein, unknown function"
820	M9	INDEL	13	PBANKA_131020	conserved <i>Plasmodium</i> protein, unknown function,
820	M9	INDEL	13	PBANKA_131030	conserved <i>Plasmodium</i> protein, unknown function,
820	M9	INDEL -2	13	PBANKA_133980	conserved <i>Plasmodium</i> protein, unknown function"
820	M9	SNP	11	PBANKA_110050	Pb-fam-1 protein"
ANKA	M9	SNP	2	PBANKA_021560	<i>Plasmodium</i> exported protein, unknown function"
ANKA	M9	SNP	4	PBANKA_040030	<i>Plasmodium</i> exported protein, unknown function"
ANKA	M9	SNP	5	PBANKA_050040	BIR protein, fragment"
ANKA	M9	INDEL	7	PBANKA_070300	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	INDEL	7	PBANKA_070580	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	INDEL	7	PBANKA_070960	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	INDEL	7	PBANKA_071350	asparagine-rich antigen, putative, fragment"
ANKA	M9	SNP - 2	7	PBANKA_072260	<i>Plasmodium</i> exported protein, unknown function"
ANKA	M9	INDEL	8	PBANKA_080020	Pb-fam-1 protein, pseudogene"
ANKA	M9	INDEL	8	PBANKA_083290	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	SNP	9	PBANKA_090020	BIR protein"
ANKA	M9	INDEL	10	PBANKA_101270	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	INDEL	12	PBANKA_124130	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	INDEL	13	PBANKA_131310	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	SNP	13	PBANKA_134050	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	INDEL	13	PBANKA_136470	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	SNP	14	PBANKA_146540	<i>Plasmodium</i> exported protein, unknown function"
ANKA	M9	INDEL	1	PBANKA_010840	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	INDEL	1	PBANKA_010850	mitochondrial ribosomal protein L19 precursor,
ANKA	M9	INDEL	1	PBANKA_011250	organic anion transporter, putative"
ANKA	M9	SNP	2	PBANKA_020030	BIR protein"
ANKA	M9	SNP - 2	3	PBANKA_031640	BIR protein, pseudogene"
ANKA	M9	INDEL	4	PBANKA_040030	<i>Plasmodium</i> exported protein, unknown function"
ANKA	M9	INDEL	4	PBANKA_040940	protein kinase, putative"
ANKA	M9	INDEL	4	PBANKA_040950	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	INDEL	5	PBANKA_051170	topoisomerase, putative"

REF STR	LINE	TYPE	CHR	GENE ID	PRODUCT
ANKA	M9	INDEL	5	PBANKA_051180	centrin-3, putative"
ANKA	M9	INDEL	6	PBANKA_062350	Pb-fam-1 protein"
ANKA	M9	INDEL	7	PBANKA_070580	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	INDEL	7	PBANKA_070590	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	INDEL	7	PBANKA_071350	asparagine-rich antigen, putative, fragment"
ANKA	M9	INDEL	7	PBANKA_072040	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	INDEL	7	PBANKA_072230	<i>Plasmodium</i> exported protein, unknown function"
ANKA	M9	SNP - 2	7	PBANKA_072240	BIR protein"
ANKA	M9	INDEL	8	PBANKA_083280	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	INDEL	8	PBANKA_083290	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	INDEL	9	PBANKA_091800	60S ribosomal protein, putative"
ANKA	M9	INDEL	9	PBANKA_093630	syntaxin, putative"
ANKA	M9	INDEL	9	PBANKA_093640	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	INDEL	10	PBANKA_101260	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	INDEL	10	PBANKA_101270	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	INDEL	10	PBANKA_101900	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	INDEL	10	PBANKA_101910	DNA-directed DNA polymerase, putative"
ANKA	M9	INDEL	10	PBANKA_102250	surface protein, putative"
ANKA	M9	INDEL	11	PBANKA_110650	rhomboid protease, putative"
ANKA	M9	INDEL	11	PBANKA_110660	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	INDEL	11	PBANKA_111520	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	INDEL	11	PBANKA_111530	metabolite/drug transporter, putative"
ANKA	M9	INDEL	11	PBANKA_111990	DNAJ protein, putative"
ANKA	M9	INDEL	11	PBANKA_112460	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	INDEL	12	PBANKA_120350	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	INDEL	12	PBANKA_120360	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	INDEL	12	PBANKA_120890	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	INDEL	12	PBANKA_122400	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	INDEL	12	PBANKA_124130	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	INDEL	13	PBANKA_130990	leucine aminopeptidase, putative"
ANKA	M9	INDEL	13	PBANKA_131310	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	INDEL	13	PBANKA_131980	serine hydroxymethyltransferase, putative"
ANKA	M9	INDEL	13	PBANKA_131990	SNARE protein, putative"

REF STR	LINE	TYPE	CHR	GENE ID	PRODUCT
ANKA	M9	INDEL	13	PBANKA_136470	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	INDEL	14	PBANKA_141010	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	INDEL - 2	14	PBANKA_142270	fe-superoxide dismutase, putative"
ANKA	M9	INDEL - 2	14	PBANKA_142280	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	INDEL	14	PBANKA_144420	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	INDEL	14	PBANKA_144440	conserved <i>Plasmodium</i> protein, unknown function"
ANKA	M9	SNP	14	PBANKA_146610	BIR protein"
<b>M8 ALONE</b>					
820	M8	SNP	2	PBANKA_020065	Pb-fam-1 protein"
820	M8	SNP	2	PBANKA_020080	<i>Plasmodium</i> exported protein, unknown function"
820	M8	SNP	2	PBANKA_020090	conserved rodent malaria protein, unknown
820	M8	SNP	2	PBANKA_021460	<i>Plasmodium</i> exported protein, unknown function,
820	M8	SNP	3	PBANKA_030060	<i>Plasmodium</i> exported protein, unknown function"
820	M8	SNP	4	PBANKA_040080	Pb-fam-1 protein, pseudogene"
820	M8	SNP	5	PBANKA_050010	Pb-fam-1 protein"
820	M8	SNP	6	PBANKA_060630	conserved <i>Plasmodium</i> protein, unknown function"
820	M8	SNP	6	PBANKA_062080	conserved <i>Plasmodium</i> protein, unknown function"
820	M8	SNP	6	PBANKA_062360	BIR protein"
820	M8	SNP	7	PBANKA_072230	<i>Plasmodium</i> exported protein, unknown function"
820	M8	SNP	12	PBANKA_124710	<i>Plasmodium</i> exported protein, unknown function"
820	M8	SNP	8	PBANKA_082080	conserved <i>Plasmodium</i> protein, unknown function"
820	M8	SNP	8	PBANKA_083700	BIR protein"
820	M8	SNP	8	PBANKA_083685	BIR protein, pseudogene"
820	M8	SNP	9	PBANKA_090020	BIR protein"
820	M8	SNP	9	PBANKA_090030	BIR protein"
820	M8	SNP	9	PBANKA_092770	conserved <i>Plasmodium</i> protein, unknown function"
820	M8	SNP	9	PBANKA_093590	conserved <i>Plasmodium</i> protein, unknown function"
820	M8	SNP	10	PBANKA_010010	Pb-fam-1 protein"
820	M8	SNP	10	PBANKA_100040	Pb-fam-1 protein"
820	M8	SNP	6	PBANKA_060010	BIR protein"
820	M8	SNP	11	PBANKA_110090	Pb-fam-1 protein"
820	M8	SNP	11	PBANKA_110130	conserved rodent malaria protein, unknown
820	M8	SNP	11	PBANKA_111840	conserved <i>Plasmodium</i> protein, unknown function"

REF STR	LINE	TYPE	CHR	GENE ID	PRODUCT
820	M8	SNP	11	PBANKA_114670	BIR protein"
820	M8	SNP	12	PBANKA_120710	<i>Plasmodium</i> exported protein, unknown function"
820	M8	SNP	12	PBANKA_122300	conserved <i>Plasmodium</i> protein, unknown function"
820	M8	SNP	12	PBANKA_124670	BIR protein"
820	M8	SNP	13	PBANKA_134050	conserved <i>Plasmodium</i> protein, unknown function"
820	M8	SNP	14	PBANKA_146580	BIR protein"
820	M8	SNP	B	PBANKA_000130	BIR protein, fragment"
820	M8	SNP	B	PBANKA_000180	<i>Plasmodium</i> exported protein, unknown function"
820	M8	SNP	B	PBANKA_000380	BIR protein"
820	M8	SNP	B	PBANKA_000390	BIR protein"
820	M8	SNP	B	PBANKA_000470	BIR protein, pseudogene"
820	M8	SNP	B	PBANKA_000540	BIR protein"
820	M8	SNP	2	PBANKA_020030	BIR protein"
820	M8	SNP	9	PBANKA_090070	BIR protein"
ANKA	M8	SNP	1	PBANKA_010010	Pb-fam-1 protein" -
ANKA	M8	SNP	1	PBANKA_010580	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	1	PBANKA_010990	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	1	PBANKA_011260	Pb-fam protein
ANKA	M8	SNP	2	PBANKA_020050	BIR protein
ANKA	M8	SNP	2	PBANKA_020060	Pb-fam-1 protein" -
ANKA	M8	SNP	2	PBANKA_021000	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	2	PBANKA_021490	Pb-fam-1 protein" -
ANKA	M8	SNP	3	PBANKA_030030	BIR protein" -
ANKA	M8	SNP	3	PBANKA_030400	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	3	PBANKA_031620	conserved rodent malaria protein
ANKA	M8	SNP	4	PBANKA_040070	Pb-fam-1 protein
ANKA	M8	SNP	4	PBANKA_040080	Pb-fam-1 protein
ANKA	M8	SNP	4	PBANKA_040400	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	4	PBANKA_041560	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	5	PBANKA_050010	Pb-fam-1 protein"
ANKA	M8	SNP	5	PBANKA_050030	conserved rodent malaria protein
ANKA	M8	SNP	5	PBANKA_051020	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	6	PBANKA_060020	BIR protein" -
ANKA	M8	SNP	6	PBANKA_060370	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	6	PBANKA_061260	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	6	PBANKA_062360	BIR protein" -
ANKA	M8	SNP	6	PBANKA_062380	BIR protein
ANKA	M8	SNP	7	PBANKA_070440	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	7	PBANKA_070580	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	7	PBANKA_070760	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	7	PBANKA_070960	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	7	PBANKA_071070	conserved <i>Plasmodium</i> protein

REF STR	LINE	TYPE	CHR	GENE ID	PRODUCT
ANKA	M8	SNP	7	PBANKA_071180	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	7	PBANKA_071950	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	7	PBANKA_072010	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	7	PBANKA_072200	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	7	PBANKA_072220	BIR protein
ANKA	M8	SNP	8	PBANKA_080010	BIR protein" -
ANKA	M8	SNP	8	PBANKA_080020	Pb-fam-1 protein
ANKA	M8	SNP	8	PBANKA_080200	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	8	PBANKA_081550	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	8	PBANKA_082080	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	8	PBANKA_082260	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	8	PBANKA_082280	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	8	PBANKA_083150	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	8	PBANKA_083685	BIR protein
ANKA	M8	SNP	9	PBANKA_090010	conserved rodent malaria protein
ANKA	M8	SNP	9	PBANKA_090030	BIR protein" -
ANKA	M8	SNP	9	PBANKA_092770	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	9	PBANKA_094400	BIR protein" -
ANKA	M8	SNP	10	PBANKA_100690	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	10	PBANKA_101990	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	10	PBANKA_102330	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	10	PBANKA_103600	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	11	PBANKA_110070	Pb-fam-1 protein" -
ANKA	M8	SNP	11	PBANKA_110075	Pb-fam-1 protein" -
ANKA	M8	SNP	11	PBANKA_110210	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	11	PBANKA_110230	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	11	PBANKA_111820	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	11	PBANKA_113900	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	11	PBANKA_114030	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	12	PBANKA_122460	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	12	PBANKA_122610	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	12	PBANKA_124030	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	12	PBANKA_124510	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	12	PBANKA_124680	BIR protein" -
ANKA	M8	SNP	13	PBANKA_130030	Pb-fam-1 protein
ANKA	M8	SNP	13	PBANKA_131410	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	13	PBANKA_131420	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	13	PBANKA_132010	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	13	PBANKA_132730	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	13	PBANKA_136120	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	13	PBANKA_136565	Pb-fam-1 protein
ANKA	M8	SNP	14	PBANKA_142460	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	14	PBANKA_144490	conserved <i>Plasmodium</i> protein
ANKA	M8	SNP	14	PBANKA_144660	conserved <i>Plasmodium</i> protein



REF STR	LINE	TYPE	CHR	GENE ID	PRODUCT
ANKA	M8	SNP	14	PBANKA_146580	BIR protein" -
ANKA	M8	SNP	14	PBANKA_146590	BIR protein" -
ANKA	M8	SNP	B	PBANKA_000170	BIR protein" -
ANKA	M8	SNP	B	PBANKA_000200	BIR protein" -
ANKA	M8	SNP	B	PBANKA_000210	Pb-fam protein
ANKA	M8	SNP	B	PBANKA_000250	BIR protein" -
ANKA	M8	SNP	B	PBANKA_000290	BIR protein" -
ANKA	M8	SNP	B	PBANKA_000370	conserved rodent malaria protein
ANKA	M8	SNP	B	PBANKA_000410	BIR protein
ANKA	M8	SNP	B	PBANKA_000490	conserved rodent malaria protein
ANKA	M8	SNP	B	PBANKA_000580	BIR protein" -

### Shade codes to the table:

Light Grey: Mutations exclusively involving UTRs

Dark Grey: Important mutations but also in the parent lines..

### Text codes to the table:

Ref Str: reference strain

Chr: Chromosome number

**Table A12 – A comprehensive list of selected prioritised mutations grouped according to their co-occurrence in the studied GNP lines**

REF STR	LINE	TYPE	CHR	GENE ID	POS	R BASE	M BASE	AA CHANGE	PRODUCT
COMMON IN ALL GNPs									
820	233	SNP	14	PBANKA_143750	1371537	C	T	R2263STOP	Transcription factor with AP2 domains
820	M7	INDEL	14	PBANKA_143750	1369916	AA	A	FSHIFT	Transcription factor with AP2 domains
820	M9	SNP	14	PBANKA_143750	1367390	T	A	Y880STOP	Transcription factor with AP2 domains
ANKA	M8	INDEL	14	PBANKA_143750	1369436	TA	A	FSHIFT	Transcription factor with AP2 domains
ANKA	233	SNP	14	PBANKA_143750	1372335	C	T	R2263STOP	Transcription factor with AP2 domains
ANKA	M7	INDEL	14	PBANKA_143750	1370714	AA	A	FSHIFT	Transcription factor with AP2 domains
ANKA	M9	SNP	14	PBANKA_143750	1368188	T	A	Y880STOP	Transcription factor with AP2 domains
ANKA	K173	INDEL	14	PBANKA_143750	1369423	CATATA	CATA	FSHIFT	Transcription factor with AP2 domains
ANKA	K173	INDEL	9	PBANKA_093910	1446117	t	tT	829US	transcription factor with AP2 domain(s),
ANKA	K173	INDEL	9	PBANKA_093910	1446332	aatatatatat atatatatata tatatat	aATatatatatata tatatatatatata at	614US	transcription factor with AP2 domain(s),
ANKA	233	INDEL	9	PBANKA_093910	1446814	t	tGT	132US	transcription factor with AP2 domain(s),
ANKA	M7	INDEL	9	PBANKA_093910	1446814	t	tGT	132US	transcription factor with AP2 domain(s),
ANKA	M9	INDEL	9	PBANKA_093910	1446814	t	tGT	132US	transcription factor with AP2 domain(s),
ANKA	M9	INDEL	9	PBANKA_093910	1446501	catatatatat atatatat	catatatatatatat at	442US	transcription factor with AP2 domain(s),
ANKA	234	INDEL	9	PBANKA_093910	1446814	ta	tTTa	132US	transcription factor with AP2 domain(s),
ANKA	820	INDEL	9	PBANKA_093910	1446813	tt	tTGt	131US	transcription factor with AP2 domain(s),
COMMON IN M7 AND M9									
820	M7	INDEL	12	PBANKA_123350	1259303	TG	TTG	FSHIFT	RNA binding protein, putative
20	M9	INDEL	12	PBANKA_123350	1259303	TGTT	TTGTT,T	FSHIFT	RNA binding protein, putative
ANKA	M7	SNP	14	PBANKA_146500	2408385	T	A	S672T	PYST-C2 homologue, putative"
ANKA	M9	SNP	14	PBANKA_146500	2408385	T	A	S672T	PYST-C2 homologue, putative"
820	M8	SNP	14	PBANKA_146500					PYST-C2 homologue, putative"

REF STR	LINE	TYPE	CHR	GENE ID	POS	R BASE	M BASE	AA CHANGE	PRODUCT
ANKA	234	SNP	14	PBANKA_146500	2408385	T	A	S672T	PYST-C2 homologue, putative"
ANKA	M7	INDEL	8	PBANKA_082560	986910	cttttttttttt	cTTTTTTTTt ttttttttt,c TTTTTTTTt tttttttttt	346DS	splicing factor 3a, putative"
ANKA	M9	INDEL	8	PBANKA_082560	986922	t	tTTTTTTT	358DS	splicing factor 3a, putative"
ANKA	234	INDEL	8	PBANKA_082560	986910	cttttttttttt	cTTTTTTTT ttttttttttt, cTTTTTTTt ttttttttt,cT TTTTTTtttt ttttttt		splicing factor 3a, putative"
COMMON IN 233 AND K173									
ANKA	233	SNP	1	PBANKA_011210	453650	T	C	F1823S	transcription factor with AP2 domain(s),
ANKA	K173	SNP	1	PBANKA_011210	453650	T	C	F1823S	transcription factor with AP2 domain(s),
ANKA	234	SNP	1	PBANKA_011210	453650	T	C	F1823S	transcription factor with AP2 domain(s),
ANKA	233	SNP	5	PBANKA_051920	701552	G	A	G131D	CCAT-binding transcription factor-like protein,
ANKA	234	SNP	5	PBANKA_051920	701552	G	A	G131D	CCAT-binding transcription factor-like protein,
ANKA	233	SNP	12	PBANKA_122210	819068	T	G	Y489D	regulator of chromosome condensation
ANKA	234	SNP	12	PBANKA_122210	819068	T	G	Y489D	regulator of chromosome condensation protein,
ANKA	233	SNP	14	PBANKA_141570	591556	G	A	M2487I	transcription factor with AP2 domain(s),
ANKA	K173	SNP	14	PBANKA_141570	591556	G	A	M2487I	transcription factor with AP2 domain(s),
ANKA	K173	SNP	14	PBANKA_141570	586073	G	T	A660S	transcription factor with AP2 domain(s),
ANKA	234	SNP	14	PBANKA_141570	591556	G	A	M2487I	transcription factor with AP2 domain(s),
233 ALONE									
820	233	SNP	10	PBANKA_102190	836265	T	A	V40E	RNA polymerase subunit, putative"
820	233	INDEL	11	PBANKA_110870	314827	ACG	AACG	FSHIFT	chromosome condensation protein putative
ANKA	233	INDEL	8	PBANKA_080330	176812	t	tT	FSHIFT	RNA helicase, putative"

REF STR	LINE	TYPE	CHR	GENE ID	POS	R BASE	M BASE	AA CHANGE	PRODUCT
ANKA	233	INDEL	11	PBANKA_112110	766876	tg	tTg	FSHIFT	radical SAM protein, putative"
ANKA	233	dsSNP	1	PBANKA_011210	?	?	?	?	transcription factor with AP2 domain(s),
ANKA	234	dsSNP	1	PBANKA_011210	?	?	?	?	transcription factor with AP2 domain(s),
ANKA	233	INDEL	3	PBANKA_031580	545510	ttat	tTat	884US	Ribosome associated memb. protein RAMP4,
ANKA	233	INDEL	14	PBANKA_145050	1909733	atgt	aTgt	545US	RNA-processing protein, putative"
820	M7	INDEL	2	PBANKA_020950	333828	AG	AAG	FSHIFT	ATP dependent RNA helicase, putative"
M7 ALONE									
820	M7	INDEL	7	PBANKA_071830	617574	T	TT	FSHIFT	DNA replication licensing factor, putative"
ANKA	M7	INDEL	7	PBANKA_071830	621042	t	Tt	FSHIFT	DNA replication licensing factor, putative"
ANKA	M7	INDEL	12	PBANKA_123350	1264140	tg	tTg	FSHIFT	RNA-binding protein, putative"
ANKA	M7	INDEL	8	PBANKA_080800	389107	catatatatat atatatatata tatatatatat atatatatata ta	catatatatatatat atatatatatatata tatatatatatata	115DS	cdc2-related protein kinase 4"
M9 ALONE									
820	M9	INDEL	6	PBANKA_061010	388494	A	AA	FSHIFT	lysine-specific histone demethylase 1, putative"
ANKA	M9	INDEL	8	PBANKA_081930	759012	ata	a	454DS	eukaryotic translation initiation factor 3
ANKA	M9	INDEL	11	PBANKA_112450	889515	catatatatat atatatatata tatatatatat atatat	catatatatatatat atatatatatatata tatatat	174DS	poly(A) polymerase PAP, putative"
ANKA	M9	INDEL	12	PBANKA_120900	321486	agt	aTAgt	210DS	eukaryotic translation initiation factor 2,
M8 ALONE									
ANKA	M8	SNP	1	PBANKA_010030	NA	NA	NA	NA	phospholipase
ANKA	M8	SNP	1	PBANKA_010040	NA	NA	NA	NA	rhopty protein
ANKA	M8	SNP	2	PBANKA_021520	NA	NA	NA	NA	reticulocyte-binding protein
ANKA	M8	SNP	2	PBANKA_021530	NA	NA	NA	NA	reticulocyte-binding protein

REF STR	LINE	TYPE	CHR	GENE ID	POS	R BASE	M BASE	AA CHANGE	PRODUCT
820	M8	SNP	3	PBANKA_030600	NA	NA	NA	NA	6-cysteine protein"
820	M8	SNP	3	PBANKA_031270	NA	NA	NA	NA	DEAD/DEAH helicase, putative
ANKA	M8	SNP	3	PBANKA_031660	NA	NA	NA	NA	reticulocyte-binding protein
820	M8	SNP	3	PBANKA_031660	NA	NA	NA	NA	reticulocyte-binding protein, putative"
ANKA	M8	SNP	4	PBANKA_041600	NA	NA	NA	NA	RhopH3
820	M8	SNP	4	PBANKA_041600	NA	NA	NA	NA	RhopH3
ANKA	M8	SNP	5	PBANKA_050100	NA	NA	NA	NA	reticulocyte binding protein
820	M8	SNP	5	PBANKA_050100	NA	NA	NA	NA	reticulocyte binding protein
ANKA	M8	SNP	5	PBANKA_051900	NA	NA	NA	NA	S-antigen
ANKA	M8	SNP	6	PBANKA_060240	NA	NA	NA	NA	transporter
ANKA	M8	SNP	6	PBANKA_060950	NA	NA	NA	NA	kinesin-related protein
ANKA	M8	SNP	6	PBANKA_061030	NA	NA	NA	NA	polyubiquitin
ANKA	M8	SNP	7	PBANKA_070100	NA	NA	NA	NA	tryptophan-rich antigen
ANKA	M8	SNP	7	PBANKA_071350	NA	NA	NA	NA	asparagine-rich antigen
ANKA	M8	SNP	7	PBANKA_071370	NA	NA	NA	NA	GTPase
ANKA	M8	SNP	8	PBANKA_080320	NA	NA	NA	NA	inositol phosphatase
ANKA	M8	SNP	8	PBANKA_081060	NA	NA	NA	NA	exoribonuclease
ANKA	M8	SNP	8	PBANKA_081490	NA	NA	NA	NA	arginyl-tRNA synthetase
820	M8	SNP	8	PBANKA_081700	NA	NA	NA	NA	Sugar transporter, putative
ANKA	M8	SNP	8	PBANKA_082600	NA	NA	NA	NA	methyltransferase
ANKA	M8	SNP	8	PBANKA_083330	NA	NA	NA	NA	zinc finger protein
ANKA	M8	SNP	9	PBANKA_090080	NA	NA	NA	NA	reticulocyte binding protein
ANKA	M8	SNP	9	PBANKA_090090	NA	NA	NA	NA	reticulocyte binding protein
ANKA	M8	SNP	9	PBANKA_090130	NA	NA	NA	NA	merozoite adhesive erythrocytic binding protein
820	M8	SNP	9	PBANKA_090130	NA	NA	NA	NA	merozoite adhesive erythrocytic binding protein
ANKA	M8	SNP	10	PBANKA_100010	NA	NA	NA	NA	reticulocyte binding protein

REF STR	LINE	TYPE	CHR	GENE ID	POS	R BASE	M BASE	AA CHANGE	PRODUCT
ANKA	M8	SNP	10	PBANKA_102410	NA	NA	NA	NA	ATP-dependent RNA helicase
ANKA	M8	SNP	10	PBANKA_102800	NA	NA	NA	NA	ubiquitin C-terminal hydrolase
ANKA	M8	SNP	10	PBANKA_102820	NA	NA	NA	NA	protein kinase
ANKA	M8	SNP	10	PBANKA_103640	NA	NA	NA	NA	BOP1-like protein
820	M8	SNP	10	PBANKA_103640	NA	NA	NA	NA	BOP1-like protein
ANKA	M8	SNP	11	PBANKA_110810	NA	NA	NA	NA	guanidine nucleotide exchange factor
ANKA	M8	SNP	11	PBANKA_110870	NA	NA	NA	NA	chromosome condensation protein
ANKA	M8	SNP	11	PBANKA_111660	NA	NA	NA	NA	alpha adaptin-like protein
ANKA	M8	SNP	11	PBANKA_112350	NA	NA	NA	NA	SNF2 helicase
ANKA	M8	SNP	11	PBANKA_113330	NA	NA	NA	NA	elongation factor 1 alpha
ANKA	M8	SNP	11	PBANKA_113340	NA	NA	NA	NA	elongation factor 1 alpha
ANKA	M8	SNP	12	PBANKA_120620	NA	NA	NA	NA	histone deacetylase
ANKA	M8	SNP	12	PBANKA_122850	NA	NA	NA	NA	ubiquitin-like protease 1 homolog
ANKA	M8	SNP	12	PBANKA_124020	NA	NA	NA	NA	G-protein coupled receptor
ANKA	M8	SNP	13	PBANKA_130650	NA	NA	NA	NA	TRAP-like protein
820	M8	SNP	13	PBANKA_131570	NA	NA	NA	NA	rhopty neck protein 2
ANKA	M8	SNP	13	PBANKA_133970	NA	NA	NA	NA	DEAD box helicase
ANKA	M8	SNP	14	PBANKA_143290	NA	NA	NA	NA	GTP-binding protein
ANKA	M8	SNP	14	PBANKA_143400	NA	NA	NA	NA	phosphate translocator
ANKA	M8	SNP	14	PBANKA_143820	NA	NA	NA	NA	DNA GyrAse a-subunit
ANKA	M8	SNP	14	PBANKA_144820	NA	NA	NA	NA	asparagine/aspartate rich protein
ANKA	M8	SNP	bin	PBANKA_000160	NA	NA	NA	NA	reticulocyte binding protein
ANKA	M8	SNP	bin	PBANKA_000220	NA	NA	NA	NA	reticulocyte binding protein
ANKA	M8	SNP	bin	PBANKA_000280	NA	NA	NA	NA	reticulocyte binding protein
ANKA	M8	SNP	bin	PBANKA_000450	NA	NA	NA	NA	reticulocyte-binding protein
ANKA	M8	SNP	bin	PBANKA_000510	NA	NA	NA	NA	reticulocyte binding protein

REF STR	LINE	TYPE	CHR	GENE ID	POS	R BASE	M BASE	AA CHANGE	PRODUCT
820	M8	SNP	bin	PBANKA_000220	NA	NA	NA	NA	reticulocyte binding protein
820	M8	SNP	bin	PBANKA_000280	NA	NA	NA	NA	reticulocyte binding protein
820	M8	SNP	bin	PBANKA_000450	NA	NA	NA	NA	reticulocyte binding protein
820	M8	SNP	bin	PBANKA_000510	NA	NA	NA	NA	reticulocyte binding protein

### Colour codes to the table:

Green: Mutations of prime importance

Orange: Other potentially relevant mutations

Grey: Mutations exclusively involving UTRs

Dark orange: Important mutations but also in the parent lines..

### Text codes to the table:

Ref Str: reference strain

Chr: Chromosome number

Pos: position of the mutation (base number)

R base: base in the reference genome

M base: base in the mutant

AA change: Amino acid change

US: upstream

DS: downstream

FSHIFT: frameshift

NOSHIFT: No frameshift

dsSNP: synonymous SNP

Table A13: Summary of all the oligonucleotides used in the study (fonts shaded in grey indicate the restriction enzyme recognition sites)

NAME	SEQUENCE	T <sub>m</sub> (°C)	RESTRICTION SITE	DESCRIPTION
GU0274	GGG <b>gaagctt</b> GTCTGTACATAAGTGAATATAC	64.2	HindIII	Forward primer, flanking pb000676.02.0 upstream
GU0275	CCC <b>gggccc</b> CAATGGTGTGCAAATAAG	69.5	Apal	Reverse primer, flanking pb000676.02.0 upstream
GU0276	GGG <b>gatatc</b> GCTGAGTGGAGCACATTGC	69.5	EcoRV	Forward primer, flanking pb000676.02.0 downstream
GU0277	CCC <b>gaattc</b> CCGAACGTAACAATACAAATTCCC	68.2	EcoRI	Reverse primer, flanking pb000676.02.0 downstream
GU0287	CATTTCCCATATTACAATTTTAGC	54.2		Forward Primer; checking the integration of SMC
GU0288	GAAAATACCCAAAGCGATAG	53.2		Reverse Primer; checking the integration of SMC
GU0587	GTGAACAAATAAAAGATAAAAG	43.8		230p 5'int forward
GU0588	GTCCAACATTTATGAATC	42.7		pbdhfr 3'UTR reverse
GU0589	GAAGTAGTGTGTAGCGTATTC	50.4		3'UTR pb48/45 forward
GU0590	GGATTCAAATCAATATTGCATC	48.1		230p 3'int reverse
GU0980	GGG <b>gaagctt</b> GTATTCTTCTCTTTTTTATGAAGGATACAAAATCGC	58.2	HindIII	Forward HA1 PBANKA141570 ko
GU0981	CCC <b>gggccc</b> CCACAATATGATTAATATTAATTTGTTATAGGGCGTGAATG	58.1	Apal	Reverse HA1 PBANKA141570 ko
GU0982	CCC <b>gggccc</b> CCCACGGTTATTCATTTTTTATTTCCACAATATG	57.4	Apal	Reverse HA1 (mix) PBANKA141570 ko
GU0983	GGG <b>ctcgag</b> GCATATATAAGCCAGATGGAAATGTAATCTACCTCAAAAAAAC	60.1	XhoI	Forward HA2 (mix) PBANKA141570 ko
GU0984	GGG <b>ctcgag</b> CTACCTCAAAAAAACTTAACCTTTTTCATTTTTTGG	55.5	XhoI	Forward HA2 PBANKA141570 ko
GU0985	CCC <b>gaattc</b> CAAGTTTAAAAAGATGTCCTAAATAGCTATGTTAAATCCG	57.8	EcoRI	Reverse HA2 PBANKA141570 ko
GU0986	GGG <b>gaagctt</b> GTTAGCTTTATTTTTTCATGATTGTAATGTATAATGCC	56.3	HindIII	Forward for complementation of PBANKA141570
GU0987	GGG <b>gaagctt</b> GCACAATATCATATCTACAGTTGTGTGCGTTTTATAGG	60	HindIII	Forward HA1 PBANKA091860 ko
GU0988	CCC <b>gggccc</b> GAAAAACTATAGCATTTGTAAAGAAGCAAACAAATTATATCC	57.8	Apal	Reverse HA1 PBANKA091860 ko
GU0989	GGG <b>ctcgag</b> GTATTGAAAAATATATTAACACATTTAGCTATTATTATGAGGC	56.1	XhoI	Forward HA2 PBANKA091860 ko
GU0990	CCC <b>gaattc</b> CATTGAAAATTATTAATAAATTATGTTTTCCCTTTCCTATGC	57.1	EcoRI	Reverse HA2 PBANKA091860 ko
GU1004	CGTTTTTTTGACGCTTTAATTTTCGTATTTGACGTTTATTG	60		5' Int check Api2 141570 ko
GU1005	GTTTTTAATTTGTTCTTCCAAGTGTGCGTCTACAG	61		3' Int check Api2 141570 ko
GU1006	GCATACTTATATAATTTCTGCATTGCATATATGCCTC	60		5' Int check Kinase 091860 ko
GU1007	CACACATAAATATATGTATATAAAAAAGGGGGGTGTATGC	62		3' Int check Kinase 091860 ko
GU1008	GGTGAAGTAATAAAAAGCATATATATTTTGAATCATCC	59		Forward for complementation of PBANKA091860
GU1009	GGG <b>gggccc</b> GTATTCTTCTCTTTTTTATGAAGGATACAAAATCGC	58.2	Apal	Forward HA1 PBANKA141570 ko
GU1010	CCC <b>ccg</b> gggCCACAATATGATTAATATTAATTTGTTATAGGGCGTGAATG	58.1	SacII	Reverse HA1 PBANKA141570 ko



GU1011	CCCccgaggCCCACGGTTATTCATTTTTATTTCCACAATATG	57.4	SacII	Reverse HA1 (mix) PBANKA141570 ko
GU1012	GGGccgaggCCAATCCTAAGGAATATAAAAAATTCTACACAAATTGAG	59	SacII	EXT FW GFP Tag PBANKA_141570
GU1013	CCCggatccCATTTCCATCTGGCTTATATATGCAGTATTATTCG	60	BamHI	EXT REV GFP Tag PBANKA_141570
GU1014	CATACAATtctagaTTATCGAAAATGAGGTCATTAATAATC	61	XbaI	INT FW GFP Tag PBANKA_141570
GU1015	CCTCATTTTCGATAAtctagaATTGTATGTTTATTTTTTCAT	59	XbaI	INT REV GFP Tag PBANKA_141570
GU1016	GGGccgaggGAAGAATTTTACATACTGGACAAATTAATCAACTAC	58	SacII	EXT FW GFP Tag PBANKA_091860
GU1017	CCCggatccTTTTTTTGATTAACTTTACTAACTCCTAAAATTTTTTAAAAAAG	57	BamHI	EXT REV GFP Tag PBANKA_091860
GU1018	GCCCTAAGCAATATCtctagaTATAATAATAAAAAAATGC	60	XbaI	INT FW GFP Tag PBANKA_091860
GU1019	TTTTTTTATTATTATAtctagaGATATTGCTTAGGGCAGAAG	60	XbaI	INT REV GFP Tag PBANKA_091860
GU1028	GTGAAGTTCAAATATGTGAAAAACAATAAATGAATTTAGC	59		FW outside 230p in 820
GU1029	GTGACTTCTAGTGAATCGCTACCATAAGTATC	62		RE outside 230p in 820
GU1030	ATGAGTAGATCTTCTAAGAACGTCATCAAGG	59		FW RFP in 820
GU1031	CAAGAACAAGTGGTGTCTACCTTCAGTACG	62		RE RFP in 820
GU1032	GTGAGCAAGGGCGAGGAGCTG	60		FW eGFP in 820 (mid)
GU1033	CTCGGCATGGACGAGCTGTACAAG	61		FW eGFP in 820 (end)
GU1034	CAGCAGGACCATGTGATCGCG	58		RE eGFP in 820
GU1054	GGGgaagcttGGTGCAATATATATGTATTTATTCATATATATACAC	60.6	HindIII	Forward HA1 PBANKA143750 ko
GU1055	CCCggggcccGTTTATTACAAATACTTCTAAATCAATATCACTAC	60.1	Apal	Reverse HA1 PBANKA143750 ko
GU1056	CCCggggcccGTTTGAATATCAGAGATCATTTTTGTTTATTAC	60.1	Apal	Reverse HA1 (mix) PBANKA143750 ko
GU1057	GGGgaagcttGTTATCTATTTTACTTCTAAAAACATGTATATATTTAATTTG	60.7	HindIII	Forward for complementation PBANKA143750 ko
GU1058	GGGctcgagCGCAATTATTATTGCATATAATAATAATAATAAC	59.7	XhoI	Forward HA2 PBANKA143750 ko
GU1059	CCCgaattcGTCACAATATGTATAAAAACGTTAATAATTTCTAAAATTG	61.2	EcoRI	Reverse HA2 PBANKA143750 ko
GU1060	GGGctcgagGAGTGTAGTAGCAATAATAATCGCC	61.6	XhoI	Forward HA2 (mix) PBANKA143750 ko
GU1061	GGGgaagcttCTTTTACATTTTCTATATTTCTATTTCAATTGG	60.1	HindIII	Forward HA1 PBANKA011210 ko
GU1062	CCCggggcccGTAACAATTATCAATCATATATACATATAGAAAAAAC	61	Apal	Reverse HA1 PBANKA011210 ko
GU1063	CCCggggcccCAATTTTCCGATTTAGAATATTTATCCATTTTAGC	61.3	Apal	Reverse HA1 (mix) PBANKA011210 ko
GU1064	GGGggggcccCCTCAAATTATATGTGTATTAATATCCAATAAAGG	61.5	Apal	Forward for complementation PBANKA011210 ko
GU1065	GGGctcgagGAATATAATCATAATTAATACTATTTATTTGCAATGTTTG	60.2	XhoI	Forward HA2 PBANKA011210 ko
GU1066	CCCgaattcGGATAAAGTTATTTATTAATAATAATAACGATTTTAGTG	60.4	EcoRI	Reverse HA2 PBANKA011210 ko
GU1067	GGGctcgagGTAGAATCGAAAGATGAATGATAAAATAAAAG	59.2	XhoI	Forward HA2 (mix) PBANKA011210 ko
GU1068	CGCCTATAACAAATTAATTAATCATATTGTGG	62.4		ORF PBANKA_141570 FW
GU1069	CCAAAAATGAAAAAGTTAAGTTTTTTTGAGGTAG	61.3		ORF PBANKA_141570 RE
GU1070	GTAGTGATATTGATTAGAAGTATTTGTAATAAAC	60.1		ORF PBANKA_143750 FW
GU1071	GTTATTATTATTATTATATGCAATAAATAATTGGCG	59.7		ORF PBANKA_143750 RE

GU1072	GTTTTTCTATATGTGTATATATGATTGATAATTGTTAC	61		ORF PBANKA_011210 FW
GU1073	CAACATTGCAAATAAATAGTTAATATTATGATTATATTC	60.2		ORF PBANKA_011210 RE
GU1074	GTTTGCTCTTTACAAATGCTATAGTTTTTC	60.2		ORF PBANKA_091860 FW
GU1075	GCCTCATAATAATAGCTAAATGTGTTAATATATTTTTCAATAC	63.7		ORF PBANKA_091860 RE
GU1076	CATATTTAAAGCTATGATAAGTAGAGTAAGCAAC	62.2		5' INT CHECK PBANKA_143750
GU1077	GTCTTGCTTGTTTCAGGTAAAATAATATAAAATGG	62.2		3' INT CHECK PBANKA_143750
GU1078	GAACATTAAATTCGGGGGTAAAGAAAATTTG	61.8		5' INT CHECK PBANKA_011210
GU1079	CATATTAATGACCGGATCAGATTACTATTTTAAATC	62.6		3' INT CHECK PBANKA_011210
GU1080	GGGgaattcGGGAAATAATAGTGACAGTAATATTTGTAATAATAC	61.5	EcoRI	FW SXO HA PBANKA_141570
GU1081	CCCctcgagGATATTTATTATATTTAACATTTTGGGTATTTATATTATCATC	60.9	XhoI	RE SXO HA PBANKA_141570
GU1082	GGGgaattcGAAACGAATGGTTGTTGTTTCATGTCAAAC	62.4	EcoRI	FW SXO HAPBANKA_011210
GU1083	CCCctcgagGAATCATATCTTTTTTCTTTTACACCTTCC	60.5	XhoI	RE SXO HA PBANKA_011210
GU1084	GGGgaattcGTAGCAATTATGATTTAAAAATCAATAACGAACATAG	61.7	EcoRI	FW SXO HA PBANKA_143750
GU1085	CCCctcgagCGGAAATATGCTATAATAAATTATTATTACAATAATTC	60.2	XhoI	RE SXO HA PBANKA_143750
GU1086	GAAATAAATAT tccggaGAAAAGTT TACATAATAACAATGATC	63.3	Kpn2I	INT FW SXO HA PBANKA_143750
GU1087	GTAAACTTTTctccggaATATTTATTTTCATTTGCACATATTAATG	65.8	Kpn2I	INT REV SXO HA PBANKA_143750
GU1125	CCCccgaggCTGTTGAGACATTTGGTATAGTTGTTTTTC	61	SacII	Reverse for HA1" PBANKA_141570
GU1126	GGGctcgagGATACAAAAAATGGATTACGTAATACAAAAAATATGG	61.7	XhoI	Forward for HA2" PBANKA_141570
GU1127	CCCgggcccCAGTCGAATTAACATCCATCAATTGAC	60.4	apaI	Reverse for HA1" PBANKA_143750
GU1128	GGGctcgagCATATAAAGAGCGTCGCTATAACAAGAAAC	62.7	XhoI	Forward for HA2" PBANKA_143750
GU1129	CCCgggcccCGGTTGTATATTTTCAGAATATTCATTATAGTTTG	61.3	apaI	Reverse for HA1" PBANKA_011210
GU1130	GGGctcgagCAACCATTTTAAAAATAATAATTTAGAGATGAATGATAAC	61.2	XhoI	Forward for HA2" PBANKA_011210
GU1139	TTYAGGGTTYAGGGTTYAGGGTTYAGGGTTYAGGGTTYAGGG	76.5		Telomere probe for CHEF gels
GU1168	GGGccgaggCCAAAAAATGTTGTTATAGAAAATAGTACTCAAATTAC	61.9	SacII	PBANKA_143750 GFP tag EXT FW
GU1169	CCCgggatccCGTACTACACTCATCATGTAGCACGCTG	68.1	BamHI	PBANKA_143750 GFP tag EXT RE
GU1170	GCCCAAGGGAtctagaTGAGTAATAATTT AGAAAATAGTTTATG	65	XbaI	PBANKA_143750 GFP tag INT FW
GU1171	CTAAATTATTACTCActagaTCCCTTGGGCAAATGGATTAGTG	67.3	XbaI	PBANKA_143750 GFP tag INT RE
GU1172	GGGccgaggCATTTAATGGATTTCAAAAATATACAAATATAAATAATGG	61.2	SacII	PBANKA_011210 GFP tag EXT FW
GU1173	CCCgggatccCATCTTTCGATTCTACTTTATTTCTATAATCGAAAATGG	64.2	BamHI	PBANKA_011210 GFP tag EXT RE
GU1182	GGGctcgagCGATGAAGGTAATAAAAGTAGAAATATAAAAAACAAC	61.7	XhoI	PBANKA_143750 SXO Complementation EXT FW
GU1183	CCCgatatcTTATTATTCGCTACTACACTCATCATGTAG	61.3	EcoRV	PBANKA_143750 SXO Complementation EXT RE
GU1184	GACAAAAATATATTAACAAATAAAGctccggaCATAGACAATC	61.7	Kpn2I	PBANKA_143750 SXO Complementation INT FW
GU1185	TTAATTCATTTTGATTGTCTATGtccggaGCTTTATTTG	59.5	Kpn2I	PBANKA_143750 SXO Complementation INT RE
GU1186	CCCgatatcGTCTTGCTTGTTTCAGGTAAAATAATATAAAATGG	62.2	EcoRV	PBANKA_143750 SXO Comp EXT_RE with homoUTR

GU1187	CCCgatatcGATCCGCATATCGAAATGATGCTATC	61.6	EcoRV	PBANKA_143750 SXO Comp EXT_RE with exoUTR
GU1231	GGGtctagaCGATGAAGGTAATAAAAGTAGAAATATAAAAAACAAC	59.5	XbaI	PBANKA_143750 SXO pL31 Comp. EXT FW (NEW)
GU1232	CCCggatccTTATTATTCGCTACTACACTCATCATGTAG	62.2	BamHI	PBANKA_143750 SXO pL31 Comp. EXT RE (NEW)
GU1233	GACAAAAATATATTAACAAATAAAGCggccggccCATAGACAATC	60.2	FseI	PBANKA_143750 SXO pL31 Comp INT FW (NEW)
GU1234	TTAATTCATTTTGATTGTCTATGggccggccGCTTTATTG	61.3	FseI	PBANKA_143750 SXO pL31 Comp INT RE (NEW)
GU1235	CCCggatccGTCTTGCTTGTCAGGTAAAATAATATAAAATGG	61.3	BamHI	PBANKA_143750 SXO pL31 Comp EXT RE with homoUTR (NEW)
GU1236	GAAAGTCTGATGAATTTTATAAAATGAGGG	60.4		Sequencing oligo for m8/K173 mutation in 820 PBANKA_143750
GU1237	CTCCTGTCGACACAAATAATGATAATTATG	61		Sequencing oligo for m7 mutation in 820 PBANKA_143750
GU1238	GGAAAGATTCAACGAGAGGACATTG	59.8		Sequencing oligo for 233 mutation in 820 PBANKA_143750
GU1239	CTGGTTTGTTACTAATCCAATATGTCC	61.3		Q1R - Rev for Q1 amplification PBANKA_143750
GU1240	CATTATTGTCCCATCATTATTTCTATTATATC	61		Q2R - Rev for Q2 amplification PBANKA_143750
GU1241	GATTATAAAATAGGCTATATTAATGACGAAAATG	59.8		Q3F - FW for Q3 amplification PBANKA_143750
GU1242	GCCCACCATTTCTGTATTTTAAAATTTCTTC	62.2		Q3R - Rev for Q3 amplification PBANKA_143750
GU1287	GTATAATGTAATGTAACGAAAACATAAGTTATTTTAAC	60.8		3' Int check sense oligo for PBANKA_143750 Repair/ko
GU1315	CCTCTATGAAGGTCAAGGGAG	59.8		PBANKA_052270 (Alpha-tubulin II) RE RTPCR oligo
GU1317	CAGGTGGGCATTCAAAAGCTATTAC	61.3		PBANKA_030610 (P230) RE RTPCR oligo (5' set)
GU1319	CCCTGAAATGCGGAATTTTAAAAGCTG	61.9		PBANKA_030610 (P230) RE RTPCR oligo (3' set)
GU1322	CTTTCCGACGTCGTTCCG	61		PBANKA_135970 (P47) Sense RTPCR oligo
GU1325	CTGAGTATATAAAATATATTTAACAAAAGTGAGC	60.1		PBANKA_051490 (P28) Sense RTPCR oligo
GU1327	CTTCAATTATACTATTGTTCAATTGGCTCTTC	61.8		PBANKA_051500 (P25) Sense RTPCR oligo
GU1338	GGGccggcgcGATTGATAATTCCTGCAGCCC	60.3	SacII	Pbeef1aa sense in pL0035 with SacII site
GU1385	GGGggcgccgcCGATGAAGGTAATAAAAGTAGAAATATAAAAAACAAC	59.5	NotI	PBANKA_143750 SXO pL31 CompEXT FW_CORRECTED
GU1386	CCCtctagaTTTCGCTACTACACTCATCATGTAGC	61.3	XbaI	PBANKA_143750 SXO pL31 Comp EXT RECORRECTED
GU1387	GACAAAAATATATTAACAAATAAAGCgaagcttCATAGACAATC	61.7	HindIII	PBANKA_143750 SXO pL31 Compl INT FW CORRECTED
GU1388	TTAATTCATTTTGATTGTCTATGaagcttGCTTTATTG	59.5	HindIII	PBANKA_143750 SXO pL31 Comp INT RE CORRECTED
GU1411	GCTTAATATATCTTCATAAAACATATTATTATTATGGC	60.8		PBANKA_143750 REV oligo for promoter repair
GU1412	GGGaagcttGGGAAAATGCGATGTGGAAATAGTTG	61.6	HindIII	PBANKA_143750 part orf with mutations ko FW HA1

GU1413	CCCccgaggGCATATTTAATTCATTTTGATTGTCTATGGC	60.2	SacII	PBANKA_143750 part orf with mutations ko RE HA1
GU1414	GGGctcgagGAAAATAATTATTGTGTTAAAGACATTTCCGAC	60.8	XhoI	PBANKA_143750 part orf with mutations ko FW HA2
GU1415	CCCgaattcCTGCCTTTCAAACCTTTTCTGG	60.6	EcoRI	PBANKA_143750 part orf with mutations ko RE HA2
GU1416	GGGagcttGAAAATATGAATAATGAAGAAAATTTAAAAATACAGAAAATG	60.9	HindIII	PBANKA_143750 part orf with DBD ko FW HA1
GU1417	CCCccgaggCATATTTAAACGTGAATGATTTGTGGGG	60.7	SacII	PBANKA_143750 part orf with DBD ko RE HA1
GU1446	AATATTATAATAGTCGTAGCCATCAATGTGTACACATGGTAATATAGATTTTCGTTTATATT	69.4		Target motif (PBANKA_143750/PFL1085w) for EMSA
GU1447	AATATTATAATAGTCGTAGCCATCAATAAAAAAAATGGTAATATAGATTTTCGTTTATATT	66.8		No motif (negative control) for EMSA
GU1448	AATATTATAATAGTCGTAGCCATCAATATATAAAAAATGGTAATATAGATTTTCGTTTATATT	66.8		Mutated motif (neg. control; G/C replaced with A)
GU1449	AATATTATAATAGTCGTAGCCATCAATTGCATGCAATGGTAATATAGATTTTCGTTTATATT	69.4		Another ApiAP2 motif (negative control; PBANKA_132980/PF14_0633) for EMSA
GU1497	GAAAAACAATAATTCTGAAAATATTATTTCTACACATC	60.6		PBANKA_143750 SXO complementation FW
GU1498	GAAAATGGAAAATTGCTATGTCCATTCAAG	61.3		PBANKA_143750 SXO complementation RE
GU1516	CATATATACATATAACCATTTTTATTCTATTGGGTC	61.3		PBANKA_143750 5' repair FW (NS & SM removal)
GU1517	CCCCAAAAGCAAAGAAGACATTTATTTGG	61		PBANKA_143750 5' repair RE (NS and SM removal)
GU1625	AATATTATAATAGTCGTAGCCATCAATGTGTACACATGGTAATATAGATTTTCGTTTATATT	69.4		Target motif (PBANKA_143750/PFL1085w) for EMSA
GU1626	AATATTATAATAGTCGTAGCCATCAATAAAAAAAATGGTAATATAGATTTTCGTTTATATT	66.8		No motif (negative control) for EMSA
GU1627	AATATTATAATAGTCGTAGCCATCAATATATAAAAAATGGTAATATAGATTTTCGTTTATATT	66.8		Mutated motif (neg. control; G/C replaced with A)
GU1628	AATATTATAATAGTCGTAGCCATCAATTGCATGCAATGGTAATATAGATTTTCGTTTATATT	69.4		Another ApiAP2 motif (negative control; PBANKA_132980/PF14_0633) for EMSA
GU1822	AATATAAACGAAAATCTATATTACCATGTGTACACATTGATGGCTACGACTATTATAATATT	69.4		REVERSE Target motif (PBANKA_143750/PFL1085w)
GU1823	AATATAAACGAAAATCTATATTACCATTTTTTTTTATTGATGGCTACGACTATTATAATATT	66.8		REVERSE No motif (negative control) for EMSA
GU1824	AATATAAACGAAAATCTATATTACCATTTTTTATATATTGATGGCTACGACTATTATAATATT	66.8		REVERSE Mutated motif (neg control; G/C replaced with A)
GU1825	AATATAAACGAAAATCTATATTACCATGTCATGCAATTGATGGCTACGACTATTATAATATT	69.4		
GU1885	CGCTAAAAAGAATTAAGGAACGCTTTC	60.2		5'IC FW SXO RepTag (from genome) AP2-G
GU1886	CAACTCCAGTGAAAAGTTCTTCTCC	61.3		5'IC RE SXO RepTag (from GFP) AP2-G
GU1887	GGGccgaggCATTGTACTAACACAAATTCCTATATTTTTC	60.8	SacII	FW PROMOTER+ORF AP2-G PBANKA_143750
GU1888	CCCaaagcttGTTGTTGTAGTGGCACTCGATTTTTC	61.9	HindIII	RE PROMOTER+ORF AP2-G PBANKA_143750
GU2068	TATGTTAAACATTGGCAAATTTATAGTACACCTAATAGTATCATTCTATTGCATCATT			PBANKA_143750 UP1F
GU2069	TATGTTAAACATTGGCAAATTTATAGTAAACCTAATAGTATCATTCTATTGCATCATT			PBANKA_143750 UP1F MUT1
GU2070	TATGTTAAACATTGGCAAATTTATAAAAAAACCTAATAGTATCATTCTATTGCATCATT			PBANKA_143750 UP1F MUT2
GU2071	TAATATACTGTATGTATATATGCATGCGTACTTATGCATTATAAATTGTCTACACAATTT			PBANKA_143750 UP2F
GU2072	TAATATACTGTATGTATATATGCATGCATACTTATGCATTATAAATTGTCTACACAATTT			PBANKA_143750 UP2F MUT1 F
GU2073	TAATATACTGTATGTATATATGCATAAAAAATTATGCATTATAAATTGTCTACACAATTT			PBANKA_143750 UP2F MUT2 F

GU2074	ATATTTTAATCTGGTTCCTTTTTCTGCGTACTTATCCAGAATGAATGTTTTAATTCATTT			PBANKA_103430F
GU2075	ATATTTTAATCTGGTTCCTTTTTCTGCGTACTTATCCAGAATGAATGTTTTAATTCATTT			PBANKA_103430F MUT1F
GU2076	ATATTTTAATCTGGTTCCTTTTTCTAAAAAATTATCCAGAATGAATGTTTTAATTCATTT			PBANKA_103430F MUT2F
GU2077	TCATTTGATCAATTTCTACAACGTGAGTACGCGCTATCAAAAATGTAAAAAAGATTTTTA			PBANKA_106086F
GU2078	TCATTTGATCAATTTCTACAACGTGAATACGCGCTATCAAAAATGTAAAAAAGATTTTTA			PBANKA_106086F MUT1F
GU2079	TCATTTGATCAATTTCTACAACGTAAAAAAGCTATCAAAAATGTAAAAAAGATTTTTA			PBANKA_106086F MUT2F
GU2080	GAAAAAATAATAATAAAGTATAAATGTGTACATCATAGTTTTATTCTAAATTCATTCCTG			PBANKA_141480F
GU2081	GAAAAAATAATAATAAAGTATAAATGTATACATCATAGTTTTATTCTAAATTCATTCCTG			PBANKA_141480F MUT1F
GU2082	GAAAAAATAATAATAAAGTATAAATAAAAAAATCATAGTTTTATTCTAAATTCATTCCTG			PBANKA_141480F MUT2F
GU2083	TTTCTTGTTTAGTGTTATAATAACGTGTACACACAACACATAAAAAATGCGCACCTATAT			PBANKA_081070F
GU2084	TTTCTTGTTTAGTGTTATAATAACGTGTAACACACAACACATAAAAAATGCGCACCTATAT			PBANKA_081070F MUT1F
GU2085	TTTCTTGTTTAGTGTTATAATAACAAAAAACAACACATAAAAAATGCGCACCTATAT			PBANKA_081070F MUT2F
GU2225	ATGGCTAACGCAAAAGCAAAGCC	60.6		HSP70 orf FW PBANKA_071190
GU2226	TTAATCAACTTCTTCAACAGTTGGTCC	60.4		HSP70 orf RE PBANKA_071190
GU2227	GGATTTTTGGAATTCTTGTGAACCAC	60.4		HSP70 orf RE (middle) PBANKA_071190
GU2242	TTTGTGGTGCACTTCATTGC	55.2		QCR1 for 143750_COMP_DOWN
GU2243	GTATATTATTTATGGTAAAATGTCGGTG	57.8		QCR2 for 143750_COMP_DOWN
GU2244	CATACTAGCCATTTTATGTG	51.1		GW1 for 143750_COMP_DOWN
GU2245	CTTTGGTGACAGATACTAC	52.4		GW2 for 143750_COMP_DOWN
GU2246	CACATTGAATTAATAAATGTGTGTGG	56.9		Genomic integration check 143750_COMP_DOWN

# A14 – List of chemicals and reagents used

## 1. LB (Luria Bertani) culture medium/agar:

1.0% peptone	5.0	(gram per 500 ml)
0.5% yeast extract	2.5	
1.0% NaCl	5.0	

Bring the components in solution and autoclave for 15 minutes at 120°C. To enrich the medium, the amount of Tryptone and Yeast Extract can be doubled. For preparing LB agar, add 1.5% agar prior to autoclaving.

## 2. SOB culture medium:

2.0% peptone	10.0	(gram per 500 ml)
0.5% yeast extract	2.5	
10 mM NaCl	0.29	
2.5 mM KCl	0.09	
10 mM MgCl <sub>2</sub> .6H <sub>2</sub> O	1.02	
10 mM MgSO <sub>4</sub> .7H <sub>2</sub> O	1.23	

Bring the components in solution and autoclave 15 minutes at 120°C. One can choose to make 100 X stock solutions of NaCl, KCl, MgCl<sub>2</sub>, and MgSO<sub>4</sub>.

## 3. SOC culture medium:

SOB culture medium plus 20 mM glucose.

After autoclaving and cooling down of the SOB medium, add a filter-sterile solution of glucose to a final concentration of 20 mM.

## 4. Denaturing buffer:

0.5 M NaOH	10.0	(gram per 500 ml)
1.5 M NaCl	43.8	

## 5. 20 X SSC stock solution:

3.0 M NaCl	175.3	(gram per 500 ml)
0.3 M Na <sub>3</sub> citrate.2H <sub>2</sub> O	88.2	

Bring the components in solution and autoclave 15 minutes at 120°C.

## 6. TNE buffer:

10 mM Tris pH 8.0  
5 mM EDTA pH 8.0  
100 mM NaCl

Combine 1 ml of 1 M Tris (pH 8.0) with 1 ml of 0.5 M EDTA (pH 8.0) and 10 ml of 1 M NaCl. Add demineralised water up to 100 ml. If required, autoclave for 15 minutes at 120°C.

## 7. 10 X TAE electrophoresis buffer:

400 mM Tris	387.6	(gram per 8 litre)
Acetic acid	-	
10 mM EDTA	-	

Dissolve Tris in ~ 3 litres of demineralised water; add 160 ml of 0.5 M EDTA (pH 8.0) and 91.5 ml of a 100% acetic acid solution. Adjust the volume up to 8 litres with demineralised water.

8. 5 X TBE electrophoresis buffer:

445 mM Tris	540	(gram per 10 litre)
445 mM Boric acid	275	
10 mM EDTA	-	

Dissolve Tris and Boric acid in ~ 3 litres of demineralised water, add 200 ml of 0.5 M EDTA (pH 8.0) and adjust the volume up to 10 litres with demineralised water.

9. TE:

10 mM Tris-Cl, pH 8.0  
1 mM EDTA, pH 8.0

Combine 10 ml 1 M Tris-Cl, pH 8.0, with 2 ml of 0.5 M EDTA, pH 8.0. Add demineralised water up to 1 litre and autoclave for 15 minutes at 120°C.

10. Parasite lysis buffer for RNA extraction:

10 mM Tris-Cl, pH 8.0  
0.4 M NaCl  
1% SDS (v/v)

11. Hepes buffer:

10 mM Hepes, pH 7.5

Prepare the buffer with DEPC treated water and autoclave for 15 minutes at 120°C.

12. 10 X RNA Gel Running Buffer:

0.2 M MOPS  
50 mM NaAc  
10 mM EDTA

Dissolve and autoclave for 15 minutes at 120°C.

13. RNA denaturing buffer:

50% (v/v) Formamide  
2.2 M Formaldehyde  
1 X Running buffer

Combine 500 µl demineralised Formamide with 179 µl of 37% formaldehyde solution (12.3 M) and 100 µl 10 X running buffer. Add demineralised water up to 1 ml.

14. RNA loading buffer (5X):

30% (w/v) Ficoll  
1 mM EDTA, pH 8.0  
0.25% (w/v) Bromo Phenol Blue (runs with ~ 0.5 Kb)  
0.25% (w/v) Xylene Cyanol FF (runs with ~ 4 Kb)

15. 5 X dATP-mix:

5 X React buffer 2 (Gibco BRL) or Sure/Cut buffer L or M (Roche)  
 2.5 mM dCTP  
 2.5 mM dGTP  
 2.5 mM dTTP

Combine 50 µl of 10 X React buffer 2 (or Sure/Cut buffer L or M) with 2.5 µl of each nucleotide (100 mM solutions). Add demineralised water up to 100 µl. Make 5 µl aliquots and store at -20°C

16. Denhardt's solution (100 X):

2% (w/v) Ficoll  
 2% (w/v) polyvinylpyrrolidone  
 2% (w/v) BSA

Dissolve the components in demineralised water. Make 50 ml aliquots in Falcon tubes and store at -20°C. Filter sterilisation is an option but not necessary.

17. Hybridisation buffer:

6 X SSC  
 5 X Denhardt's  
 0.5% (v/v) SDS  
 0.1% (v/v) disodium pyrophosphate  
 15 µg/ml tRNA

For 100 ml: take about 50 ml of demineralised water and add 30 ml of 20 X SSC, 5 ml of 100 X Denhardt's solution, 5 ml of a 10% (w/v) SDS solution, 1 ml of a 10% (w/v) disodium pyrophosphate solution, and 150 µl of a 10 mg/ml tRNA solution. Add demineralised water up to 100 ml. Before use, place at 60°C.

18. PBS (phosphate buffered saline) normal:

The 10x stock is the commercial available PBS from Roche (Cat. nr: 1666 789).

Components of the 10x PBS: 0.01 M  $\text{KH}_2\text{PO}_4$ , 1.37 M NaCl, and 0.027 M KCl, pH 7.0

Working solution:

- Dilute the stock 10 x with demineralised water
- Adjust the pH to 7.2 with 1 M HCl solution
- Autoclave for 20 minutes

19. PBS (phosphate buffered saline) 'rich':

20mM Hepes, 20 mM Glucose, 4mM  $\text{NaHCO}_3$ , 0.1% BSA

Working solution:

- Dilute the stock PBS solution 10 times (see above) and add the above mentioned components
- Adjust the pH to 7.25 with 5 M NaOH
- Sterilise the solution by passing it through a 0.2 µm filter
- Store the solution at 4°C



20. Giemsa-solution and staining buffer:

For dilution of the stock Giemsa solution (Merck Cat. nr. 1.09204.0500) we use the Sörensen staining buffer.

*Sörensen staining buffer:*

- $\text{KH}_2\text{PO}_4$  2,541 g per 5 liter
- $\text{Na}_2\text{HPO}_4 \cdot 2\text{H}_2\text{O}$  5507 g per 5 liter
- Adjust the pH to 7.2 with NaOH

We use a 12 % Giemsa-buffer solution for staining slides for a period of 10 minutes.

21. Complete culture medium: RPMI1640 pH 7.3 containing 25% foetal calf serum:

Culture medium: RPMI1640, with L-glutamine and 25 mM HEPES, without  $\text{NaHCO}_3$  (Invitrogen).

Preparation of medium:

- Dissolve 15,89 g powder of RPMI1640 medium in 1l demineralised water. Add powder slowly under continuous stirring.
- Add 0.85 g  $\text{NaHCO}_3$
- Add 50.000 I.U. Neomycin (stock-solution of 10.000 I.U./ml; Gibco)
- Sterilise by filtration through a 0,2  $\mu\text{m}$  sieve (see below)
- Store at  $-20^\circ\text{C}$  in 100-200 ml bottles
- Immediately prior to use, add foetal calf serum (FCS) to a final concentration of 25% (v/v) to give complete culture medium.

Sterilisation of culture medium

- Fill reservoir with medium and close lid
- Connect reservoir to  $\text{N}_2$ -gas bottle and to the filter container which is placed in the flow hood
- Open gas bottle (0.5 bar)
- At the same time open the small tap on the filter container until the air has escaped.
- Open the other tap and collect the medium

Sterilisation of the filter container

- Remove the upper part of the filter container by removing the three screws
- Place the Metrical membrane (0.2  $\mu\text{m}$ ) with the shiny side downwards on the metal sieve. The blue side of the metal sieve should be facing upwards
- Place the polypropylene sieve on the Metrical membrane
- Place the glassfiber filter on top and close the filter container
- Autoclave for 20 min. Taps and in-and outlets covered with aluminium foil.

Filters

1. Metrical Membrane filter G/A-8; 0.2  $\mu\text{m}$ , 142 mm
2. Polypropylene filter, 127 mm
3. Glassfiber filter A/E, 127 mm

22. Ookinete culture medium: RPMI1640 pH 8.0 containing 10% fetal calf serum:

- RPMI1640 medium prepared and stored as complete culture medium (see above)
- The pH is adjusted to 8.2 with 1N NaOH prior to sterilisation
- Immediately prior to use, foetal calf serum (FCS) is added at a final concentration of 10% (v/v) to give ookinete culture medium with a pH of about 8.0.

23. Nycodenz stock solution:

Nycodenz powder (Lucron Bioproduct BV) is obtained from Life Technologies. Store at room temperature. Dissolve 138g Nycodenz-powder in 500 ml Buffered Medium (see below) (density 1.15g/ml at 20°C). Autoclave for 20min at 120°C and store at 4°C

Buffered medium:

5mmol/l Tris/HCl	605.7 mg/l, pH 7.5
3mmol/l KCl	223.7 mg/l
0.3 mmol/l Ca Na <sub>2</sub> EDTA	112.3 mg/l

24. Erythrocyte-lysis buffer:

10x stock-solution:

1.5M NH <sub>4</sub> Cl	80.23 g/l
0.1M KHCO <sub>3</sub>	10.012 g/l
0.01M EDTA	3.72 g/l

## References

- Adessi C, Matton G, Ayala G, Turcatti G, Mermod JJ, Mayer P and Kawashima E (2000) Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms. *Nucleic Acids Res.* 28:E87.
- Agarwal S, Kern S, Halbert J, Przyborski JM, Baumeister S, Dandekar T, Doerig C, and Pradel G (2011) Two nucleus-localized CDK-like kinases with crucial roles for malaria parasite erythrocytic replication are involved in phosphorylation of splicing factor. *J. Cell. Biochem.* 112(2):1295-1310.
- Agnandji ST, Lell B, Soulanoudjingar SS, Fernandes JF, Abossolo BP, et al. (2011) First results of phase 3 trial of RTS,S/AS01 malaria vaccine in African children. *N Engl J Med.* 365(20):1863-75.
- Agnandji ST, Lell B, Fernandes JF, Abossolo BP, Methogo BG, Kabwende AL et al. (2012) A Phase 3 trial of RTS,S/AS01 malaria vaccine in African infants. *N Engl J Med*, 367(24):2284-2295.
- Aikawa M, Miller LH, Johnson J, and Rabbege J (1978) Erythrocyte entry by malarial parasites. A moving junction between erythrocyte and parasite. *J Cell Biol.* 77(1):72-82.
- Alano P, Roca L, Smith D, Read D, Carter R and Day K (1995) *Plasmodium falciparum*: Parasites defective in early stages of gametocytogenesis. *Exp. Parasitol.* 81:227-235.
- Alano P and Billker O (2005) Gametocytes and gametes. In *Molecular Approaches to Malaria*. Sherman, I. (ed.). Washington, DC; American Society for Microbiology Press, pp: 191-219.
- Alano P (2007) *Plasmodium falciparum* gametocytes: still many secrets of a hidden life. *Molecular Microbiology.* 66(2); 291-302.
- Albuquerque SS, Carret C, Grosso AR, Tarun AS, Peng X, Kappe SH, Prudêncio M, and Mota MM (2009) Host cell transcriptional profiling during malaria liver stage infection reveals a coordinated and sequential set of biological events. *BMC Genomics.* 10:270.
- Alonso PL, Brown G, Arevalo-Herrera M, Binka F, Chitnis C, Collins F, Doumbo OK, Greenwood B, Hall BF, Levine MM, Mendis K, Newman RD, Plowe CV, Rodriguez MH, Sinden R, Slutsker L, and Tanner M (2011) A research agenda to underpin malaria eradication. *PLoS Med*, 8:e1000406.
- Altschul SF, Wootton JC, Zaslavsky E, and Yu Y-K (2010) The construction and use of logodds substitution scores for multiple sequence alignment. *PLoS Comput Biol*, 6:e1000852.

- Aly AS, Vaughan AM and Kappe SH (2009) Malaria Parasite Development in the Mosquito and Infection of the Mammalian Host. *Annu Rev Microbiol.* 63: 195-221.
- Amino R, Thiberge S, Martin B, Celli S, Shorte S, Frischknecht F and Menard R (2006) Quantitative imaging of *Plasmodium* transmission from mosquito to mammal. *Nat. Med.*, 12; 220-224.
- Amoutzias GD, Robertson DL, Van de Peer Y and Oliver SG (2008) Choose your partners: dimerization in eukaryotic transcription factors. *Trends Biochem. Sci.* 33, 220-229.
- Andenmatten N, Egarter S, Jackson AJ, Jullien N, Herman JP, and Meissner M (2013) Conditional genome engineering in *Toxoplasma gondii* uncovers alternative invasion mechanisms. *Nat Methods* 10:125-127.
- Antinori S, Galimberti L, Milazzo L, and Corbellino M (2013) *Plasmodium knowlesi*: the emerging zoonotic malaria parasite. *Acta Trop.* 125(2):191-201.
- Arastu-Kapur S, Ponder EL, Fonović UP, Yeoh S, Yuan F, Fonović M, Grainger M, Phillips CI, Powers JC and Bogyo M (2008) Identification of proteases that regulate erythrocyte rupture by the malaria parasite *Plasmodium falciparum*. *Nat. Chem. Biol.* (4); 203-213.
- Aravind L, Iyer LM, Wellems TE and Miller LH (2003) *Plasmodium* biology: genomic gleanings. *Cell*, 115, 771-785.
- Arbeitman MN, Furlong EE, Imam F, Johnson E, Null BH, Baker BS, Krasnow MA, Scott MP, Davis RW, and White KP (2002) Gene expression during the life cycle of *Drosophila melanogaster*. *Science.* 297; 2270-2275.
- Ariey F, Witkowski B, Amaratunga C, Beghain J, Langlois AC, Khim N, Kim S, Duru V, Bouchier C, Ma L, Lim P, Leang R, Duong S, Sreng S, Suon S, Chuor CM, Bout DM, Ménard S, Rogers WO, Genton B, Fandeur T, Miotto O, Ringwald P, Le Bras J, Berry A, Barale JC, Fairhurst RM, Benoit-Vical F, Mercereau-Puijalon O, Ménard D (2014) A molecular marker of artemisinin-resistant *Plasmodium falciparum* malaria. *Nature.* 505(7481):50-5.
- Arnot DE and Gull K (1998) The *Plasmodium* cell-cycle: facts and questions. *Annals of Tropical Medicine & Parasitology* (92):4; 361-365.
- Assefa S, Keane TM, Otto TD, Newbold C, and Berriman M (2009) ABACAS: algorithm-based automatic contiguation of assembled sequences. *Bioinformatics.* 1; 25(15):1968-9.

- Babiker HA, Schneider P and Reece SE (2008) Gametocytes: insights gained during a decade of molecular monitoring. *Trends in Parasitology*, 24;11: 525-530.
- Badis G, Chan E, van Bakel H, Pena-Castillo L, Tillo D, Tsui K, Carlson C, Gossett A, Hasinoff M, Warren C, Gebbia M, Talukder S, Yang A, Mnaimneh S, Terterov D, Coburn D, Yeo AL, Yeo ZX, Clarke ND, Lieb JD, Ansari AZ, Nislow C, and Hughes TR (2008) A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. *Mol. Cell* 32; 878-887.
- Baker DA (2010) Malaria gametocytogenesis. *Molecular & Biochemical Parasitology*. 172; 57-65.
- Balaji S, Madan Babu M, Iyer LM and Aravind L (2005) Discovery of the principal specific transcription factors of Apicomplexa and their implication for the evolution of the AP2-integrase DNA binding domains. *Nucleic Acids Research*. 33(13); 3994-4006.
- Balu B and Adams JH (2007) Advancements in transfection technologies for *Plasmodium*. *Int. J Parasitol.* 37; 1-10.
- Balu B, Chauhan C, Maher SP, Shoue DA, Kissinger JC, Fraser MJ Jr, and Adams JH (2009) piggyBac is an effective tool for functional analysis of the *Plasmodium falciparum* genome. *BMC Microbiol.* 9; 83.
- Bandara LR, Girling R, and La Thangue NB (1997) Apoptosis induced in mammalian cells by small peptides that functionally antagonize the Rb-regulated E2F transcription factor. *Nat Biotechnol.* 15(9):896-901.
- Bannister AJ, and Kouzarides T (2011) Regulation of chromatin by histone modifications. *Cell Research*. 21:381-395.
- Bannister LH, Hopkins JM, Fowler RE, Krishna S, and Mitchell GH (2000) A brief illustrated guide to the ultrastructure of *Plasmodium falciparum* asexual blood stages. *Parasitol Today*. 16(10):427-33.
- Bannister LH, and Mitchell G (2003) The ins, outs and roundabouts of malaria. *Trends Parasitol.* (19):5;209-213.
- Baton LA, and Ranford-Cartwright LC (2005) Spreading the seeds of million-murdering death: metamorphoses of malaria in the mosquito. *Trends Parasitol.* 21(12):573-80.
- Baum J, Papenfuss AT, Mair GR, Janse CJ, Vlachou D, Waters AP, Cowman AF, Crabb BS, and De Koning-Ward TF (2009) Molecular genetics and comparative

- genomics reveal RNAi is not functional in malaria parasites. *Nucleic Acids Res*, 37:3788-3798.
- Beier JC, Onyango FK, Koros JK, Ramadhan M, Ogwang R, Wirtz R, Koech DK and Roberts CR (1991) Quantitation of malaria sporozoites transmitted in vitro during salivation by wild Afrotropical *Anopheles*. *Med Vet Entomol* 5:71-79.
- Bentley DR (2006) Whole-genome re-sequencing. *Curr. Opin. Genet. Dev.* 16:545-552.
- Berger MF, Philippakis AA, Qureshi AM, He FS, Estep PW, and Bulyk ML (2006) Compact, universal DNA microarrays to comprehensively determine transcriptionfactor binding site specificities. *Nat Biotechnol*, 24(11):1429-1435.
- Berger MF and Bulyk ML (2009) Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. *Nature Protocols*. 4 (3) 393-411.
- Berger SL, Kouzarides T, Shiekhata R, and Shilatifard A (2009) An operational definition of epigenetics. *Genes Dev* 23: 781-783.
- Bergman LW, Kaiser K, Fujioka H, Coppens I, Daly TM, Fox S, Matuschewski K, Nussenzweig V, and Kappe SH (2003) Myosin A tail domain interacting protein (MTIP) localizes to the inner membrane complex of *Plasmodium* sporozoites. *J Cell Sci*; 116(Pt 1):39-49.
- Berman BP, Pfeiffer BD, Laverty TR, Salzberg SL, Rubin GN, Eisen MB, and Celniker SE (2004) Computational identification of developmental enhancers: Conservation and function of transcription factor binding-site clusters in *Drosophila melanogaster* and *Drosophila pseudoobscura*. *Genome Biol* 5:R61.
- Bhasin VK and Trager W (1984) Gametocyte forming and non-gametocyte forming clones of *Plasmodium falciparum*. *American Journal of Tropical Medicine and Hygiene*, 3:534-537.
- Billker O, Shaw MK, Margos G and Sinden RE (1997) The roles of temperature, pH and mosquito factors as triggers of male and female gametogenesis of *Plasmodium berghei* in vitro. *Parasitology*. 114;1-7.
- Billker O, Lindo V, Panico M, Etienne AE, Paxton T, Dell A, Rogers M, Sinden RE and Morris HR (1998) Identification of xanthurenic acid as the putative inducer of malaria development in the mosquito. *Nature*. 392;289-292.

- Billker O, Dechamps S, Tewari R, Wenig G, Franke-Fayard B and Brinkmann V (2004) Calcium and a calcium-dependent protein kinase regulate gamete formation and mosquito transmission in a malaria parasite. *Cell*. 117;503-514.
- Birago C, Bucci A, Dore E, Frontali C and Zenobi P (1982) Mosquito infectivity is directly related to the proportion of repetitive DNA in *Plasmodium berghei*. *Molecular and Biochemical Parasitology*. 6:1-12.
- Birago C, Pace T, Picci L and Ponzi M (1994) Isolation of a distally located gene possibly correlated with gametocyte production ability. *Mem. Inst. Oswaldo Cruz*. 89:Suppl.II; 33-35.
- Bischoff E and Vaquero C (2010) In silico and biological survey of transcription-associated proteins implicated in the transcriptional machinery during the erythrocytic development of *Plasmodium falciparum*. *BMC Genomics*. 11(1); article 34.
- Blagborough AM and Sinden RE (2009) *Plasmodium berghei* HAP2 induces strong malaria transmission-blocking immunity in vivo and in vitro. *Vaccine*. 27(38):5187-94.
- Blank TA and Becker PB (1996) The Effect of Nucleosome Phasing Sequences and DNA Topology on Nucleosome Spacing. *J. Mol. Biol.*, 260:1-8.
- Bonasio R, Tu S, and Reinberg D (2010) Molecular signals of epigenetic states. *Science* 330: 612-616.
- Bopp SE, Manary MJ, Bright AT, Johnston GL, Dharia NV, Luna FL, McCormack S, Plouffe D, McNamara CW, Walker JR, Fidock DA, Denchi EL, and Winzeler EA (2013) Mitotic evolution of *Plasmodium falciparum* shows a stable core genome but recombination in antigen families. *PLoS Genet*. 9(2):e1003293.
- Borges S, Cravo P, Creasey A, Fawcett R, Modrzynska K, Rodrigues L, Martinelli A, and Hunt P (2011) Genomewide scan reveals amplification of *mdr1* as a common denominator of resistance to mefloquine, lumefantrine, and artemisinin in *Plasmodium chabaudi* malaria parasites. *Antimicrob Agents Chemother*. 55(10):4858-65.
- Boschet C, Gissot M, Briquet S, Hamid Z, Claudel-Renard and Vaquero (2004) Characterization of PfMyb1 transcription factor during erythrocytic development of 3D7 and F12 *Plasmodium falciparum* clones. *Molecular and Biochemical Parasitology*. 138(1);159-163.



- Bougdour A, Braun L, Cannella D, and Hakimi MA (2010) Chromatin modifications: implications in the regulation of gene expression in *Toxoplasma gondii*. *Cell Microbiol.* 12(4):413-23.
- Bowen B, Steinberg J, Laemmli UK and Weintraub H (1980) The detection of DNA-binding proteins by protein blotting. *Nucleic Acids Res.* 8;1-20.
- Boyle MJ, Wilson DW and Beeson JG. (2013) New approaches to studying *Plasmodium falciparum* merozoite invasion and insights into invasion biology. *Int J Parasitol.* 43(1):1-10.
- Bozdech Z, Llinas M, Pulliam BL, Wong ED, Zhu J, and DeRisi JL (2003) The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biol.* 1:E5.
- Bozdech Z, Zhu J, Joachimiak MP, Cohen FE, Pulliam B and DeRisi JL (2003) Expression profiling of the schizont and trophozoite stages of *Plasmodium falciparum* with a long oligonucleotide microarray. *Genome Biol.* 4: R9.
- Braks JA, Franke-Fayard B, Kroeze H, Janse CJ, and Waters AP (2006) Development and application of a positive-negative selectable marker system for use in reverse genetics in *Plasmodium*. *Nucleic Acids Res.* 34(5):e39.
- Braks JA, Mair GR, Franke-Fayard B, Janse CJ and Waters AP (2008) A conserved U-rich RNA region implicated in regulation of translation in *Plasmodium* female gametocytes. *Nucleic Acids Res.* 36: 1176-1186.
- Brancucci NMB, Bertschi NL, Lei Zhu L, Niederwieser I, Chin WH, Wampfler R, Freymond CL, Rottmann M, Felger I, Bozdech Z, and Voss TS (2014) Heterochromatin Protein 1 Secures Survival and Transmission of Malaria Parasites. *Cell Host & Microbe.* 16: 165-176.
- Branda CS, and Dymecki SM (2004) Talking about a revolution: the impact of site-specific recombinases on genetic analyses in mice. *Dev. Cell.* 6:7-28.
- Breman JG, Alilio MS and Mills A (2004) Conquering the intolerable burden of malaria: what's new, what's needed: a summary. *Am. J. Trop. Med. Hyg.* 71:1-15.
- Breman JG (2012) Resistance to artemisinin-based combination therapy. *Lancet Infect Dis.* 12(11):820-2.
- Brennan JD, Kent M, Dhar R, Fujioka H and Kumar N (2000) *Anopheles gambiae* salivary gland proteins as putative targets for blocking transmission of malaria parasites. *Proc Natl Acad Sci USA.* 97;13859-13864.

- Bright AT, and Winzeler EA (2013) Epidemiology: resistance mapping in malaria. *Nature*. 498(7455):446-7.
- Briquet S, Boschet C, Gissot M, Tissandie E, Sevilla E, Franetich JF, Thiery I, Hamid Z, Bourgouin C, and Vaquero C (2006) High-mobility-group box nuclear factors of *Plasmodium falciparum*. *Eukaryotic Cell*. 5(40) 672-682.
- Brockelman CR (1982) Conditions favoring gametocytogenesis in culture of *Plasmodium falciparum*. *J. Parasitology*. 29:454-458.
- Bruce MC, Alano P, Duthie S, and Carter R (1990) Commitment of the malaria parasite *Plasmodium falciparum* to sexual and asexual development. *Parasitology*. 100:191-200.
- Bruce MC, Carter RN, Nakamura K, Aikawa M and Carter R (1994) Cellular location and temporal expression of the *Plasmodium falciparum* sexual stage antigen Pfs16. *Mol Biochem Parasitol*. 65:11-22.
- Buckling A, Ranford-Cartwright LC, Miles A, and Read AF (1999) Chloroquine increases *Plasmodium falciparum* gametocytogenesis in vitro. *Parasitology*. 118:339-346.
- Buckling A and Read AF (2001) The effect of partial host immunity on the transmission of malaria parasites. *Proc R Soc Lond B Biol Sci*. 68:2325-2330.
- Bullen HE, Charnaud SC, Kalanon M, Riglar DT, Dekiwadia C, Kangwanransan N, Torii M, Tsuboi T, Baum J, Ralph SA, Cowman AF, de Koning-Ward TF, Crabb BS, and Gilson PR (2012) Biosynthesis, localization, and macromolecular arrangement of the *Plasmodium falciparum* translocon of exported proteins (PTEx). *J Biol Chem*. 287:7871-84.
- Bullen HE, Crabb BS and Gilson PR (2012) Recent insights into the export of PEXEL/HTS-motif containing proteins in *Plasmodium* parasites. *Curr Opin Microbiol*. 15:699-704.
- Burkot T, Williams JL, and Schneider I (1984) Infectivity to mosquitoes of *Plasmodium falciparum* clones grown in vitro from the same isolate. *Trans R Soc Trop Med Hyg*. 78:339-341.
- Bustamante CD, Fledel-Alon A, Williamson S, Nielsen R, Hubisz MT, Glanowski S, Tanenbaum DM, White TJ, Sninsky JJ, Hernandez RD, Civello D, Adams MD, Cargill M, and Clark AG (2005) Natural selection on protein-coding genes in the human genome. *Nature*. 20;437(7062):1153-7.

- Butler J, MacCallum I, Kleber M, Shlyakhter IA, Belmonte MK, Lander ES, Nusbaum C, and Jaffe DB (2008) ALLPATHS: de novo assembly of whole-genome shotgun microreads. *Genome Res.* 18:810-20.
- Cai Yu-Hang and Huang He (2012) Advances in the study of protein-DNA interaction. *Amino Acids.* 43:1141-1146.
- Campbell CC (2009) Malaria Control – Addressing Challenges to Ambitious Goals. *N Engl J Med.* 361:5;522-523.
- Campbell TL, De Silva EK, Olszewski KL, Elemento O, and Llinas M (2010) Identification and Genome-Wide Prediction of DNA Binding Specificities for the ApiAP2 Family of Regulators from the Malaria Parasite. *PLoS Pathog* 6(10): e1001165.
- Cao Y, Yao Z, Sarkar D, Lawrence M, Sanchez GJ, Parker MH, MacQuarrie KL, Davison J, Morgan MT, Ruzzo WL, Gentleman RC, and Tapscott SJ (2010) Genome-wide MyoD binding in skeletal muscle cells: a potential for broad cellular reprogramming. *Dev. Cell* 18, 662-674.
- Carlton JM, Angiuoli SV, Suh BB, Kooij TW, Pertea M, Silva JC, Ermolaeva MD, Allen JE, Selengut JD, Koo HL, Peterson JD, Pop M, Kosack DS, Shumway MF, Bidwell SL, Shallom SJ, van Aken SE, Riedmuller SB, Feldblyum TV, Cho JK, Quackenbush J, Sedegah M, Shoaibi A, Cummings LM, Florens L, Yates JR, Raine JD, Sinden RE, Harris MA, Cunningham DA, Preiser PR, Bergman LW, Vaidya AB, van Lin LH, Janse CJ, Waters AP, Smith HO, White OR, Salzberg SL, Venter JC, Fraser CM, Hoffman SL, Gardner MJ, and Carucci DJ (2002) Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*. *Nature.* 419:512-519.
- Carlton JM, Silva J and Hall N (2005) The Genome of Model Malaria Parasites, and Comparative Genomics. *Curr. Issues Mol. Biol.* 7:23-38.
- Carr A and Biggin MD (1999) A comparison of in vivo and in vitro DNA-binding specificities suggests a new model for homeoprotein DNA binding in *Drosophila* embryos. *EMBO J.* 18: 1598-1608.
- Carter LM, Kafack BFC, Llinas M, Mideo N, Pollitt LC and Reece SE (2013) Stress and sex in malaria parasites Why does commitment vary? *Evolution, Medicine, and Public Health.* 135-147.
- Carter R, and Miller LH (1979) Evidence for environmental modulation of gametocytogenesis in *Plasmodium falciparum* in continuous culture. *Bull World Health Organ.* 57 Suppl 1:37-52.

- Carter R. and Graves, PM. Gametocytes. In: Wernsdorfer WH and McGregor I, editor. *Malaria principles and practice of malariology*, vol. 1. Edinburgh: Churchill Livingstone; 1988. p. 253-305.
- Carter R, Graves PM, Creasey A, Byrne K, Read D, Alano P, and Fenton B (1989) *Plasmodium falciparum*: an abundant stage-specific protein expressed during early gametocyte development. *Exp Parasitol.* 69(2):140-9.
- Carvalho TG, Thiberge S, Sakamoto H, and Ménard R (2004) Conditional mutagenesis using site-specific recombination in *Plasmodium berghei*. *Proc Natl Acad Sci U S A.* 101(41):14931-6.
- Casaglia O, Dore E, Frontali C, Zenobi P and Walliker D (1985) Re-examination of earlier work on repetitive DNA and mosquito infectivity in rodent malaria. *Molecular and Biochemical Parasitology.* 16:35-42.
- Chaal BK, Gupta AP, Wastuwidyaningtyas BD, Luah YH, Bozdech Z (2010) Histone deacetylases play a major role in the transcriptional regulation of the *Plasmodium falciparum* life cycle. *PLoS Pathog*; 6:e1000737.
- Chaisson M, Pevzner P and Tang H (2004) Fragment assembly with short reads. *Bioinformatics.* 20:2067-2074.
- Chaisson M and Pevzner P (2008) Short read fragment assembly of bacterial genomes. *Genome Res.* 18:324-330.
- Chakrabarti K, Pearson M, Grate L, Sterne-Weiler T, Deans J, Donohue JP, Ares M (2007) Structural RNAs of known and unknown function identified in malaria parasites by comparative genomics and RNA analysis. *RNA.* 13:1923-1939.
- Chakravarty S, Cockburn IA, Kuk S, Overstreet MG, Sacci JB and Zavala F (2007) CD8(+) T lymphocytes protective against malaria liver stages are primed in skin-draining lymph nodes. *Nat Med.* 13:1035-1041.
- Chang HH, Falick AM, Carlton PM, Sedat JW, DeRisi JL and Marletta MA (2008) N-terminal processing of proteins exported by malaria parasites. *Mol Biochem Parasitol.* 160:107-115.
- Chatterjee S, Ngonseu E, Van Overmeir C, Correwyn A, Druilhe P, and Wéry M. (2001) Rodent malaria in the natural host--irradiated sporozoites of *Plasmodium berghei* induce liver-stage specific immune responses in the natural host *Grammomys surdaster* and protect immunized *Grammomys* against *P. berghei* sporozoite challenge. *Afr J Med Med Sci.* 30 Suppl:25-33.

- Cheeseman IH, Gomez-Escobar N, Carret CK, Ivens A, Stewart LB, Tetteh KKA and Conway DJ (2009) Gene copy number variation throughout the *Plasmodium falciparum* Genome. *BMC Genomics*. 10:353.
- Cheeseman IH, Miller BA, Nair S, Nkhoma S, Tan A, Tan JC, Al Saai S, Phyto AP, Moo CL, Lwin KM, McGready R, Ashley E, Imwong M, Stepniewska K, Yi P, Dondorp AM, Mayxay M, Newton PN, White NJ, Nosten F, Ferdig MT, and Anderson TJ (2012) A major genome region underlying artemisinin resistance in malaria. *Science*. 336(6077):79-82
- Chen K and Rajewsky N (2007) The evolution of gene regulation by transcription factors and microRNAs. *Nat Rev Genet*. 8(2):93-103.
- Chen K, McLellan MD, Ding L, Wendl MC, Kasai Y, Wilson RK and Mardis ER (2007) PolyScan: an automatic indel and SNP detection approach to the analysis of human resequencing data. *Genome Res*. 17:659-666.
- Chervitz SA, Aravind L, Sherlock G, Ball CA, Koonin EV, Dwight SS, Harris MA, Dolinski K, Mohr S, Smith T, Weng S, Cherry JM, and Botstein D (1998) Comparison of the complete protein sets of worm and yeast: orthology and divergence. *Science*. 282: 2022-2028.
- Chitnis CE and Blackman MJ (2000) Host Cell Invasion by Malaria Parasites. *Molecular Approaches to Malaria. Parasitol Today*. 16:10;411-415.
- Chookajorn T, Dzikowski R, Frank M, Li F, Jiwani AZ, Hartl DL, and Deitsch KW (2007) Epigenetic memory at malaria virulence genes. *Proc. Natl. Acad. Sci. U. S. A*. 104: 899-902.
- Clavijo CA, Mora CA and Winograd E (1998) Identification of novel membrane structures in *Plasmodium falciparum* infected erythrocytes. *Mem. Inst. Oswaldo Cruz* (93);115-120.
- Clements AC, Reid HL, Kelly GC, and Hay SI (2013) Further shrinking the malaria map: how can geospatial science help to achieve malaria elimination? *Lancet Infect Dis*. 8:709-18.
- Coleman BI and Duraisingh MT (2008) Transcriptional control and gene silencing in *Plasmodium falciparum*. *Cellular Microbiology*. 10(10): 1935-1946.
- Coleman BI, Skillman KM, Jiang RHY, Childs LM, Altenhofen LM, Ganter M, Leung Y, Goldowitz I, Kafsack BFC, Marti M, Llinas M, Buckee CO, and Duraisingh MT (2014) A *Plasmodium falciparum* Histone Deacetylase Regulates Antigenic Variation and Gametocyte Conversion. *Cell Host & Microbe*. 16: 177-186.

- Collas P (2010) The Current State of Chromatin Immunoprecipitation. *Mol Biotechnol.* 45:87-100.
- Collins CC, Das S, Wong EH, Andenmatten N, Stallmach R, Hackett F, Herman J-P, Müller S, Meissner M, and Blackman MJ (2013) Robust inducible Cre recombinase activity in the human malaria parasite *Plasmodium falciparum* enables efficient gene deletion within a single asexual erythrocytic growth cycle. *Molecular Microbiology.* 88(4), 687-701.
- Collins CR, Hackett F, Strath M, Penzo M, Withers-Martinez C, Baker DA and Blackman MJ (2013) Malaria Parasite cGMP-dependent Protein Kinase Regulates Blood Stage Merozoite Secretory Organelle Discharge and Egress. *PLoS Pathog.* 9(5): e1003344.
- Combe A, Moreira C, Ackerman S, Thiberge S, Templeton TJ, and Ménard R (2009) TREP, a novel protein necessary for gliding motility of the malaria sporozoite. *Int J Parasitol.* 39(4):489-96.
- Cornelissen AWCA (1988) Sex determination and sex differentiation in malaria parasites. *Biol. Rev.* 63: 379-394.
- Corte´s A, Crowley VM, Vaquero A, and Voss TS (2012) A View on the Role of Epigenetics in the Biology of Malaria Parasites. *PLoS Pathog* 8(12): e1002943.
- Cosma MP (2002) Ordered recruitment: gene-specific mechanism of transcription activation. *Mol. Cell.* 10: 227-236.
- Cowman AF, Galatis D and Thompson JK (1994) Selection for mefloquine resistance in *Plasmodium falciparum* is linked to amplification of the pfmdr1 gene and cross-resistance to halofantrine and quinine. *Proceedings of the National Academy of Sciences, USA.* 91:1143-1147.
- Cowman AF, Baldi DL, Healer J, Mills KE, O'Donnell RA, Reed MB, Triglia T, Wickham ME and Crabb BS (2000) Functional analysis of proteins involved in *Plasmodium falciparum* merozoite invasion of red blood cells. *FEBS Letters.* 476;84-88.
- Cowman AF and Crab BS (2006) Invasion of red blood cells by malaria parasites. *Cell.* 124:755-766.
- Cowman AF, Berry D, and Baum J (2012) The cellular and molecular basis for malaria parasite invasion of the human red blood cell. *J Cell Biol.* 198(6):961-71.
- Cox-Singh J, Davis TM, Lee KS, Shamsul SS, Matusop A, Ratnam S, Rahman HA, Conway DJ and Singh B (2008) *Plasmodium knowlesi* malaria in humans is

- widely distributed and potentially life threatening. Clin. Infect. Dis. 46:165-171.
- Crabb BS (2002) Transfection technology and the study of drug resistance in the malaria parasite *Plasmodium falciparum*. Drug Resist Updat. 5(3-4):126-30.
- Crosnier C, Bustamante LY, Bartholdson SJ, Bei AK, Theron M, Uchikawa M, Mboup S, Ndir O, Kwiatkowski DP, Duraisingh MT, Rayner JC, and Wright GJ (2011) Basigin is a receptor essential for erythrocyte invasion by *Plasmodium falciparum*. Nature. 480(7378):534-7.
- Cui F, Cole HA, Clark DJ, and Zhurkin VB (2012) Transcriptional activation of yeast genes disrupts intragenic nucleosome phasing. Nucleic Acids Res. 40(21):10753-64.
- Cui L, Fan Q, and Li J (2002) The malaria parasite *Plasmodium falciparum* encodes members of the Puf RNA-binding protein family with conserved RNA binding activity. Nucleic Acids Res. 30: 4607-4617.
- Cui L, Miao J, Furuya T, Li X, Su XZ, and Cui L (2007) PfGCN5-mediated histone H3 acetylation plays a key role in gene expression in *Plasmodium falciparum*. Eukaryot. Cell. 6: 1219-1227.
- Cui L, Fan Q, Cui L, and Miao J (2008) Histone lysine methyltransferases and demethylases in *Plasmodium falciparum*. Int. J. Parasitol. 38, 1083-1097.
- Cui L, and Miao J (2010) Chromatin-mediated epigenetic regulation in the malaria parasite *Plasmodium falciparum*. Eukaryot Cell. 9:1138-49.
- Czesny B, Goshu S, Cook JL, and Williamson KC (2009) The proteasome inhibitor epoxomicin has potent *Plasmodium falciparum* gametocytocidal activity. Antimicrob Agents Chemother. 53:4080-4085.
- Daher W, and Soldati-Favre D (2009) Mechanisms controlling glideosome function in apicomplexans. Curr Opin Microbiol. 12:408-414.
- Day KP, Karamalis F, Thompson J, Barnes DA, Peterson C, Brown H, Brown GV and Kemp DJ (1993) Genes necessary for expression of a virulence determinant and for transmission of *Plasmodium falciparum* are located on a 0.3 megabase region of chromosome 9. Proc. Natl. Acad. Sci. U. S. A., 90: 8292-8296.
- de Koning-Ward TF, Janse CJ and Waters AP (2000) The development of genetic tools for dissecting the biology of malaria parasites. Annu. Rev. Microbiol. 54;157-185.

- de Koning-Ward TF, Olivieri A, Bertuccini L, Hood A, Silvestrini F, Charvalias K, Berzosa Díaz P, Camarda G, McElwain TF, Papenfuss T, Healer J, Baldassarri L, Crabb BS, Alano P, and Ranford-Cartwright LC (2008) The role of osmiophilic bodies and Pfg377 expression in female gametocyte emergence and mosquito infectivity in the human malaria parasite *Plasmodium falciparum*. *Mol Microbiol.* 67(2):278-90.
- de Koning-Ward TF, Gilson PR, Boddey JA, Rug M, Smith BJ, Papenfuss AT, Sanders PR, Lundie RJ, Maier AG, Cowman AF and Crabb BS (2009) A newly discovered protein export machine in malaria parasites. *Nature* (459):945-9.
- De S, Lopez-Bigas N, and Teichmann SA (2008) Patterns of evolutionary constraints on genes in humans. *BMC Evol Biol.* 8:275.
- De Silva EK, Gehrke AR, Olszewski K, León I, Chahal JS, Bulyk ML, and Llinás M (2008) Specific DNA-binding by apicomplexan AP2 transcription factors. *Proc. Natl. Acad. Sci.* 105(24): 8393-8.
- Dearsly AL, Sinden RE and Self IA (1990) Sexual development in malarial parasites: gametocyte production, fertility and infectivity to the mosquito vector. *Parasitology.* 100: 359-368.
- Deitsch KW, Calderwood MS and Wellems TE (2001) Malaria. Cooperative silencing elements in var genes. *Nature.* 412: 875-876.
- Del Portillo HA, and Chitnis CE (2013) Talking to each other to initiate sexual differentiation. *Cell.* 153(5):945-7.
- Denwood MJ, Love S, Innocent GT, Matthews L, McKendrick IJ, Hillary N, Smith A, and Reid SW (2012) Quantifying the sources of variability in equine faecal egg counts: implications for improving the utility of the method. *Vet Parasitol.* 188(1-2):120-6.
- Dervan PB (2001) Molecular recognition of DNA by small molecules. *Bioorg Med Chem.* 9:2215-35.
- Dervan PB, and Edelson BS (2003) Recognition of the DNA minor groove by pyrrole-imidazole polyamides. *Curr Opin Struct Biol.* 13(3):284-99
- Dessens JT, Mendoza J, Claudianos C, Vinetz JM, Khater E, Hassard S, Ranawaka GR and Sinden RE (2001) Knockout of the rodent Malaria parasite Chitinase PbCHT1 reduces infectivity to mosquitoes. *Infect. Immun.* 69:4041-4047.
- Dessens JT, Sidén-Kiamos I, Mendoza J, Mahairaki V, Khater E, Vlachou D, Xu XJ, Kafatos FC, Louis C, Dimopoulos G, and Sinden RE (2003) SOAP, a novel



- malaria ookinete protein involved in mosquito midgut invasion and oocyst development. *Mol Microbiol.* 49(2):319-29.
- Dixon MWA, Thompson J, Gardiner DL and Trenholme KR (2008) Sex in *Plasmodium*: a sign of commitment. *Trends in Parasitology.* 24(4):168-175.
- Doerig C, Doerig C, Horrocks P, Coyle J, Carlton J, Sultan A, Arnot D, and Carter R (1995) Pfrck-1, a developmentally regulated cdc2-related protein kinase of *Plasmodium falciparum*. *Molecular and Biochemical Parasitology.* 70(1-2):167-174.
- Doerig C, Parzy D, Langsley G, Horrocks P, Carter R and Doerig CD (1996) A MAP kinase homologue from the human malaria parasite, *Plasmodium falciparum*. *Gene.* 177(1-2):1-6.
- Dohm JC, Lottaz C, Borodina T and Himmelbauer H (2007) SHARCGS, a fast and highly accurate short-read assembly algorithm for *de novo* genomic sequencing. *Genome Res.* 17:1697-1706.
- Doolan DL, Dobaño C and Baird JK (2009) Acquired Immunity to Malaria. *Clin. Microbiol. Rev.* 22(1); 13-36
- Dore E, Pace T, Picci L and Frontali C (1990) Organization of Subtelomeric Repeats in *Plasmodium berghei*. *Mol. Cell. Biol.* 10:2423-2427.
- Douradinha B, Augustijn KD, Moore SG, Ramesar J, Mota MM, Waters AP, Janse CJ, and Thompson J (2011) *Plasmodium* Cysteine Repeat Modular Proteins 3 and 4 are essential for malaria parasite transmission from the mosquito to the host. *Malar J.* 10:71.
- Dowse T and Soldati D (2004) Host cell invasion by the apicomplexans: the significance of microneme protein proteolysis. *Curr Opin Microbiol.* 7:4;388-96.
- Drakeley C, Secka I, Correa S, Greenwood BM and Targett GA (1999) Host haematological factors influencing the transmission of *Plasmodium falciparum* gametocytes to *Anopheles gambiae* s.s. mosquitoes. *Trop Med Int Health.* 4:131-138.
- van Driel R, Fransz PF and Verschure PJ (2003) The eukaryotic genome: a system regulated at different hierarchical levels. *J Cell Sci.* 116: 4067-4075.
- Duraisingh MT, Triglia T, and Cowman AF (2002) Negative selection of *Plasmodium falciparum* reveals targeted gene deletion by double crossover recombination. *Int. J. Parasitol.* 32, 81-89.

- Duraisingh MT, Voss TS, Marty AJ, Duffy MF, Good RT, Thompson JK, Freitas-Junior LH, Scherf A, Crabb BS, and Cowman AF (2005) Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in *Plasmodium falciparum*. *Cell*. 121: 13-24.
- Dyer M, and Day KP (2000). Commitment to gametocytogenesis in *Plasmodium falciparum*. *Parasitol Today*. 16:102-107.
- Dzikowski R and Deitsch KW (2009) Genetics of antigenic variation in *Plasmodium falciparum*. *Curr. Genet*. 55:103-110.
- Ecker A, Bushell ESC, Tewari R and Sinden RE (2008) Reverse genetics screen identifies six proteins important for malaria development in the mosquito. *Molecular Microbiology*. 70:1;209-220.
- Ejigiri I and Sinnis P (2009) *Plasmodium* sporozoite-host interactions from the dermis to the hepatocyte. *Current Opinion in Microbiology*. 12:401-407.
- Ekland EH and Fidock DA (2007) Advances in understanding the genetic basis of antimalarial drug resistance. *Curr Opin Microbiol*. 10(4):363-70.
- Eksi S, Haile Y, Furuya T, Ma L, Su X and Williamson KC (2005) Identification of a subtelomeric gene family expressed during the asexual-sexual stage transition in *Plasmodium falciparum*. *Mol Biochem Parasitol*. 143:90-99.
- Eksi S, Czesny B, van Gemert GJ, Sauerwein RW, Eling W and Williamson KC (2006) Malaria transmission-blocking antigen, Pfs230, mediates human red blood cell binding to exflagellating male parasites and oocyst production. *Mol Microbiol*. 61(4):991-8.
- Eksi S, Morahan BJ, Haile Y, Furuya T, Jiang H, Ali O, Xu H, Kiattibutr K, Suri A, Czesny B, Adeyemo A, Myers TG, Sattabongkot J, Su XZ and Williamson KC (2012) *Plasmodium falciparum* gametocyte development 1 (Pfgdv1) and gametocytogenesis early gene identification and commitment to sexual development. *PLoS Pathog*. 8(10):e1002964.
- Ellekvist P, Maciel J, Mlambo G, Ricke CH, Colding H, Klaerke DA, and Kumar N (2008) Critical role of a K<sup>+</sup> channel in *Plasmodium berghei* transmission revealed by targeted gene disruption. *PNAS*. 105:6398-6402.
- Elnitski L, Jin VX, Farnham PJ, and Jones SJ (2006) Locating mammalian transcription factor binding sites: a survey of computational and experimental techniques. *Genome Res*. 16(12):1455-64.
- Elowitz M and Lim WA (2010) Build life to understand it. *Nature* 468, 889-890.

- Epigentek website: [www.epigentek.com](http://www.epigentek.com). Epigentek Group Inc. (2013) Histone modification table.
- Epiphanio S, Mikolajczak SA, Gonçalves LA, Pamplona A, Portugal S, Albuquerque S, Goldberg M, Rebelo S, Anderson DG, Akinc A, Vornlocher HP, Kappe SH, Soares MP, and Mota MM (2008) Heme oxygenase-1 is an anti-inflammatory host factor that promotes murine *Plasmodium* liver infection. *Cell Host Microbe*. 3(5):331-8.
- Fakhouri WD, Ay A, Sayal R, Dresch J, Dayringer E, and Arnosti DN (2010) Deciphering a transcriptional regulatory code: modeling short-range repression in the *Drosophila* embryo. *Mol Syst Biol*. 6:341.
- Falae A, Combe A, Amaladoss A, Carvalho T, Menard R, and Bhanot P (2010) Role of *Plasmodium berghei* cGMP-dependent protein kinase in late liver stage development. *J Biol Chem*. 285(5):3282-8.
- Fan Q, Li J, Kariuki M, and Cui L (2004) Characterization of PfPuf2, member of the Puf family RNA-binding proteins from the malaria parasite *Plasmodium falciparum*. *DNA Cell Biol*. 23: 753-760.
- Feachem RG, and Sabot O (2008) A new global malaria eradication strategy. *Lancet*. 371(9624):1633-5.
- Feachem RG, and Phillips AA (2009) Malaria: 2 years in the fast lane. *Lancet*. 373(9673):1409-11.
- Feachem RG, Phillips AA and Targett GA (2009) Shrinking the Malaria Map: A prospectus on Malaria Elimination. The Global Health Group. San Francisco.
- Feachem RG, Phillips AA, Hwang J, Cotter C, Wielgosz B, Greenwood BM, Sabot O, Rodriguez MH, Abeyasinghe RR, Ghebreyesus TA, and Snow RW (2010) Shrinking the malaria map: progress and prospects. *Lancet*. 376(9752):1566-78.
- Fidock DA, Nomura T, Talley AK, Cooper RA, Dzekunov SM, Ferdig MT, Ursos LM, Sidhu AB, Naudé B, Deitsch KW, Su XZ, Wootton JC, Roepe PD, and Wellem TE (2000) Mutations in the *P. falciparum* digestive vacuole transmembrane protein PfCRT and evidence for their role in chloroquine resistance. *Mol Cell*. 6(4):861-71.
- Field JW and Shute PG (1956) The microscopic diagnosis of human malaria. II. A morphological study of the erythrocytic parasites. *Studies from the Institute for Medical Research, Federation of Malaya*. 24:251.

- Finn RD, Tate J, Mistry J, Coghill PC, Sammut SJ, Hotz HR, Ceric G, Forslund K, Eddy SR, Sonnhammer EL, and Bateman A (2008) The Pfam protein families database. *Nucleic Acids Res.* 36: D281-D288.
- Fischer JJ, Toedling J, Krueger T, Schueler M, Huber W and Sperling S (2008) Combinatorial effects of four histone modifications in transcription and differentiation. *Genomics.* 91(1):41-51.
- Florent I, Maréchal E, Gascuel O and Bréhélin L (2010) Bioinformatic strategies to provide functional clues to the unknown genes in *Plasmodium falciparum* genome. *Parasite.* 17(4): 273-283.
- Flueck C, Bartfai R, Volz J, Niederwieser I, Salcedo-Amaya AM, Alako BTF, Ehlgén F, Ralph SA, Cowman AF, Bozdech Z, Stunnenberg HG, and Voss TS (2009) *Plasmodium falciparum* heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors. *PLoS Pathog* 5: e1000569.
- Flueck C, Bartfai R, Niederwieser I, Witmer K, Alako BT, Moes S, Bozdech Z, Jenoe P, Stunnenberg HG, Voss TS (2010) A major role for the *Plasmodium falciparum* ApiAP2 protein PfSIP2 in chromosome end biology. *PLoS Pathog* 6: e1000784.
- Flueck C, and Baker DA (2014) Malaria Parasite Epigenetics: When Virulence and Romance Collide. *Cell Host & Microbe.* 16: 148-150.
- Föllner M, Bobbala D, Koka S, Huber SM, Gulbins E, and Lang F (2009) Suicide for survival--death of infected erythrocytes as a host mechanism to survive malaria. *Cell Physiol Biochem.* (3-4):133-40.
- Fonager J, Franke-Fayard BMD, Adams JH, Ramesar J, Klop O, Khan SM, Janse CJ, and Waters AP (2011) Development of the piggyBac transposable system for *Plasmodium berghei* and its application for random mutagenesis in malaria parasites *BMC Genomics.* 12:155.
- Fondon JW and Garner HR (2004) Molecular origins of rapid and continuous morphological evolution. *Proc Natl Acad Sci U S A,* 101(52):18058-18063.
- Foth BJ, Goedecke MC, and Soldati D (2006) New insights into myosin evolution and classification. *Proc Natl Acad Sci USA.* 103(10):3681-3686.
- Foth BJ, Zhang N, Mok S, Preiser PR and Bozdech Z (2008) Quantitative protein expression profiling reveals extensive post-transcriptional regulation and post-translational modifications in schizont-stage malaria parasites. *Genome Biology.* 9(12): R177.

- Francoeur AM, Gritzmacher CA, Peebles CL, Reese RT and Tan EM (1985) Synthesis of small nuclear ribonucleoprotein particles by the malarial parasite *Plasmodium falciparum*. Proc. Natl. Acad. Sci. USA. 82: 3635-3639.
- Freitas-Junior LH, Bottius E, Pirrit LA, Deitsch KW, Scheidig C, Guinet F, Nehrbass U, Wellems TE, and Scherf A (2000) Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*. Nature. 407: 1018-1022.
- Freitas-Junior LH, Hernandez-Rivas R, Ralph SA, Montiel-Condado D, Ruvalcaba-Salazar OK, Rojas-Meza AP, Mâncio-Silva L, Leal-Silvestre RJ, Gontijo AM, Shorte S, and Scherf A (2005) Telomeric heterochromatin propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites. Cell. 121: 25-36.
- Frevert U (2004) Sneaking in through the back entrance: the biology of malaria liver stages. Trends Parasitol. 20(9):417-24.
- Frevert U, Engelmann S, Zougbe de S, Stange J, Ng B, Matuschewski K, Liebes L and Yee H (2005) Intravital observation of *Plasmodium berghei* sporozoite infection of the liver. PLoS Biol. 3, e192.
- Fried M, and Crothers DM (1981) Equilibria and kinetics of lac repressor- operator interactions by polyacrylamide gel electrophoresis. Nucleic Acids Res. 9, 6505-6525.
- Frischknecht F, Baldacci P, Martin B, Zimmer C, Thiberge S, Olivo-Marin JC, Shorte SL, and Ménard R (2004) Imaging movement of malaria parasites during transmission by *Anopheles* mosquitoes. Cell Microbiol. 6(7):687-94.
- Fry CJ and Peterson CL (2001) Chromatin remodeling enzymes: who's on first? Curr. Biol. 11: R185-R197.
- Furey, TS (2012) ChIP-seq and Beyond: new and improved methodologies to detect and characterize protein-DNA interactions. Nat Rev Genet.; 13(12): 840-852.
- Furuya T, Mu J, Hayton K, Liu A, Duan J, Nkrumah L, Joy DA, Fidock DA, Fujioka H, Vaidya AB, Wellems TE, and Su XZ (2005) Disruption of a *Plasmodium falciparum* gene linked to male sexual development causes early arrest in gametocytogenesis. Proc Natl Acad Sci U S A. 102:16813-16818.
- Galas DJ and Schmitz A (1978) DNase footprinting: a simple method for the detection of protein-DNA binding specificity. Nucleic Acids Res. 5: 3157-3170.

- Ganapathi M, Palumbo MJ, Ansari SA, He Q, Tsui K, Nislow C, and Randall H (2011) Extensive role of the general regulatory factors, Abf1 and Rap1, in determining genome-wide chromatin structure in budding yeast. *Nucleic Acids Res.* 39(6): 2032-2044.
- Ganesan SM, Morrissey JM, Ke H, Painter HJ, Laroiya K, Phillips MA, Rathod PK, Mather MW, and Vaidya AB (2011) Yeast dihydroorotate dehydrogenase as a new selectable marker for *Plasmodium falciparum* transfection. *Mol Biochem Parasitol.* 177(1):29-34.
- Garcia GE, Wirtz RA and Rosenberg R (1997) Isolation of a substance from the mosquito that activates *Plasmodium* fertilization. *Mol. Biochem. Parasitol.* 88: 127-135.
- Garcia GE, Wirtz RA, Barr JR, Woolfitt A and Rosenberg R (1998) Xanthurenic Acid Induces Gametogenesis in *Plasmodium*, the Malaria Parasite. *J. Biol. Chem.* 273;12003-12005.
- Gardiner DL, Dixon MW, Spielmann T, Skinner-Adams TS, Hawthorne PL, Ortega MR, Kemp DJ and Trenholme KR (2005) Implication of a *Plasmodium falciparum* gene in the switch between asexual reproduction and gametocytogenesis. *Mol. Biochem. Parasitol.* 140:153-160.
- Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, Paulsen IT, James K, Eisen JA, Rutherford K, Salzberg SL, Craig A, Kyes S, Chan MS, Nene V, Shallom SJ, Suh B, Peterson J, Angiuoli S, Perteau M, Allen J, Selengut J, Haft D, Mather MW, Vaidya AB, Martin DM, Fairlamb AH, Fraunholz MJ, Roos DS, Ralph SA, McFadden GI, Cummings LM, Subramanian GM, Mungall C, Venter JC, Carucci DJ, Hoffman SL, Newbold C, Davis RW, Fraser CM, and Barrell B (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 419:498-511.
- Gardner MJ, Shallom SJ, Carlton JM, Salzberg SL, Nene V, Shoaibi A, Ciecko A, Lynn J, Rizzo M, Weaver B, Jarrahi B, Brenner M, Parvizi B, Tallon L, Moazzez A, Granger D, Fujii C, Hansen C, Pederson J, Feldblyum T, Peterson J, Suh B, Angiuoli S, Perteau M, Allen J, Selengut J, White O, Cummings LM, Smith HO, Adams MD, Venter JC, Carucci DJ, Hoffman SL, and Fraser CM (2002) Sequence of *Plasmodium falciparum* chromosomes 2, 10, 11 and 14. *Nature.* 419: 531-534.
- Garner MM, and Revzin A (1981) A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the

- Escherichia coli lactose operon regulatory system. *Nucleic Acids Res.* 9, 3047-3060.
- Garnham PCC (1966) *Malaria Parasites and Other Haemosporidia*. Blackwell Scientific Publications; Oxford.
- Gaskins E, Gilk S, DeVore N, Mann T, Ward G, and Beckers C (2004) Identification of the membrane receptor of a class XIV myosin in *Toxoplasma gondii*. *J Cell Biol.* 165(3): 383-93.
- Gething PW, Patil AP, Smith DL, Guerra CA, Elyazar IRF, Johnston GL, Tatem AJ and Hay SI (2011) A new world malaria map: *Plasmodium falciparum* endemicity in 2010. *Malaria Journal.* 10: 378.
- Ghosh AK, and Jacobs-Lorena M (2009) *Plasmodium* sporozoite invasion of the mosquito salivary gland. *Curr Opin Microbiol.* 12(4):394-400.
- Ghosh AK, Devenport M, Jethwaney D, Kalume DE, Pandey A, Anderson VE, Sultan AA, Kumar N and Jacobs-Lorena M (2009) Malaria parasite invasion of the mosquito salivary gland requires interaction between the *Plasmodium* TRAP and the *Anopheles* saglin proteins. *PLOS Pathog* 5(1) e1000265.
- Ghosh AK, Coppens I, Gårdsvoll H, Ploug M, and Jacobs-Lorena M (2011) *Plasmodium* ookinetes coopt mammalian plasminogen to invade the mosquito midgut. *Proc Natl Acad Sci U S A.* 108(41):17153-8.
- Gilson PR and Crabb BS (2009) Morphology and kinetics of the three distinct phases of red blood cell invasion by *Plasmodium falciparum* merozoites. *Int J Parasitol.* 39(1):91-6.
- Gissot M, Refour P, Briquet S, Boschet C, Coupé S, Mazier D, and Vaquero C (2004) Transcriptome of 3D7 and its gametocyte-less derivative F12 *Plasmodium falciparum* clones during erythrocytic development using a gene-specific microarray assigned to gene regulation, cell cycle and transcription factors. *Gene* 341:267-77.
- Gissot M, Briquet S, Refour P, Boschet C and Vaquero C (2005) PfMyb1, a *Plasmodium falciparum* transcription factor, is required for intra-erythrocytic growth and controls key genes for cell cycle regulation. *Journal of Molecular Biology.* 346(1): 29-42.
- Gissot M, Ting L, Daly T, Bergman L, Sinnis P and Kim K (2008) High mobility group protein HMGB2 is a critical regulator of *Plasmodium* oocyst development. *Journal of Biological Chemistry.* 283(25): 17030-17038.

- Global Malaria Action Plan (2008) Key figures on Malaria. World Health Organization. RBM/SEC/2009/ISS.2
- Glushakova S, Yin D, Li T and Zimmerberg J (2005) Membrane transformation during malaria parasite release from human red blood cells. *Curr. Biol.* (15):1645-1650.
- Godiska R, Mead D, Dhodda V, Wu C, Hochstein R, Karsi A, Usdin K, Entezam A, and Ravin N (2010) Linear plasmid vector for cloning of repetitive or unstable sequences in *Escherichia coli*. *Nucleic Acids Res.* 38, e88
- Goldberg DE, Janse CJ, Cowman AF, and Waters AP (2011) Has the time come for us to complement our malaria parasites? *Trends Parasitol.* 27(1):1-2.
- Grant PA (2001) A tale of histone modifications. *Genome Biology.* 2;(4):0003.1-0003.6
- Graves PM, Carter R and McNeill KM (1984) Gametocyte production in cloned lines of *Plasmodium falciparum*. *American Journal of Tropical Medicine and Hygiene.* 33:1045-1050.
- Greenwood B (2009) Can malaria be eliminated? Research, education and capacity development in resource-poor settings - a Festschrift for Professor M.E. Molyneux OBE. *Transactions of the Royal Society of Tropical Medicine and Hygiene.* 103;Suppl. 1:S2-S5.
- Greenwood B and Owusu-Agyei S (2012) Epidemiology. Malaria in the post-genome era. *Science.* 5;338(6103):49-50.
- Grüning C, Heiber A, Kruse F, Flemming S, Franci G, Colombo SF, Fasana E, Schoeler H, Borgese N, Stunnenberg HG, Przyborski JM, Gilberger TW, Spielmann T (2012) Uncovering common principles in protein export of malaria parasites. *Cell Host Microbe.* 12:717-29.
- Gu M and Lima CD (2005) Processing the message: structural insights into capping and decapping mRNA. *Curr. Opin. Struct. Biol.* 15: 99-106.
- Gueirard P, Tavares J, Thiberge S, Bernex F, Ishino T, Milon G, Franke-Fayard B, Janse CJ, Ménard R, and Amino R (2010) Development of the malaria parasite in the skin of the mammalian host. *Proc Natl Acad Sci USA.* 107:18640-18645.
- Guerra CA, Howes RE, Patil AP, Gething PW, Van Boeckel TP, Temperley WH, Kabaria CW, Tatem AJ, Manh BH, Elyazar IR, Baird JK, Snow RW, and Hay SI (2010) The international limits and population at risk of *Plasmodium vivax* transmission in 2009. *PLoS Negl Trop Dis.* 4(8):e774.



- Guinet F, Dvorak JA, Fujioka H, Keister DB, Muratova O, Kaslow DC, Aikawa M, Vaidya AB and Wellems TE (1996) A Developmental defect in *Plasmodium falciparum* male gametogenesis. *J. Cell Biol.* 135:269-278.
- Haase S and de Koning-Ward TF (2010) New insights into protein export in malaria parasites. *Cell Microbiol.* 12:580-7.
- Hahn S, and Young ET (2011) Transcriptional regulation in *Saccharomyces cerevisiae*: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. *Genetics.* 189: 705-736.
- Hakimi MA and Deitsch KW (2007) Epigenetics in Apicomplexa: Control of gene expression during cell cycle progression, differentiation and antigenic variation. *Curr Opin Microbiol.* 10:357-362.
- Halder K, Murphy SC, Milner DA Jr. and Taylor TE (2007) Malaria: Mechanisms of erythrocytic infection and pathological correlates of severe disease. *Annu. Rev. Pathol. Mech. Dis.* 2;217-49.
- Hall BS, Daramola OO, Barden G, and Targett GA (1997) Modulation of protein kinase C activity in *Plasmodium falciparum*-infected erythrocytes. *Blood.* 89(5):1770-8.
- Hall N and Carlton J (2005) Comparative genomics of malaria parasites. *Curr Opin Genet Dev.* 5(6):609-13.
- Hall N, Karras M, Raine JD, Carlton JM, Kooij TWA, Berriman M, Florens L, Janssen CS, Pain A, Christophides GK, James K, Rutherford K, Harris B, Harris D, Churcher C, Quail MA, Ormond D, Doggett J, Trueman HE, Mendoza J, Bidwell SL, Rajandream M-A, Carucci DJ, Yates III JR, Kafatos FC, Janse CJ, Barrell B, Turner CMR, Waters AP, and Sinden RE (2005a) A Comprehensive Survey of the *Plasmodium* Life Cycle by Genomic, Transcriptomic, and Proteomic Analyses. *Science.* 307:82-86.
- Hamilton WD (1967) Extraordinary sex ratios. *Science.* 156: 477-488.
- Han YS, Thompson J, Kafatos FC and Barillas-Mury C (2000) Molecular interactions between *Anopheles stephensi* midgut cells and *Plasmodium berghei*: the time bomb theory of ookinete invasion of mosquitoes. *EMBO Journal.* 19:6030- 6040.
- Han YS, Thompson J, Kafatos FC, and Barillas-Mury C (2001) Molecular interactions between *Anopheles stephensi* midgut cells and *Plasmodium berghei*: the time bomb theory of ookinete invasion of mosquitoes. *EMBO J.* 19(22):6030-40. Erratum in: *EMBO J* 20(6):1483.

- Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne JB, Reynolds DB, Yoo J, Jennings EG, Zeitlinger J, Pokholok DK, Kellis M, Rolfe PA, Takusagawa KT, Lander ES, Gifford DK, Fraenkel E, and Young RA (2004) Transcriptional regulatory code of a eukaryotic genome. *Nature*. 431: 99-104.
- Harrison T, Samuel BU, Akompong T, Hamm H, Mohandas N, Lomasney JW, Haldar K (2003) Erythrocyte G protein-coupled receptor signaling in malarial infection. *Science*. 301:1734-6.
- Hartley PD and Madhani HD (2009) Mechanisms that specify promoter nucleosome location and identity. *Cell* 137: 445-458.
- Hasan S (2003) Prediction and analysis of nucleosome positioning in genomic sequences. Chapter 1 - A General Introduction to Nucleosomes and Nucleosome Positioning. A dissertation submitted to the University of Cambridge for the degree of Doctor of Philosophy, April 2003. Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK
- Hasty P, Rivera-Perez J and Bradley A (1991) The length of homology required for gene targeting in embryonic stem cells. *Mol. Cell. Biol.* 11;5586-5591.
- Hawking F, Wilson ME, and Gammage K (1971) Evidence for cyclic development and short-lived maturity in the gametocytes of *Plasmodium falciparum*. *Trans R Soc Trop Med Hyg.* 65:549-59.
- Heiber A, Kruse F, Pick C, Grüning C, Flemming S, Oberli A, Schoeler H, Retzlaff S, Mesén-Ramírez P, Hiss JA, Kadekoppala M, Hecht L, Holder AA, Gilberger TW, and Spielmann T (2013) Identification of new PNEPs indicates a substantial non-PEXEL exportome and underpins common features in *Plasmodium falciparum* protein export. *PLoS Pathog.* 9:e1003546.
- Heintzelman MB, and Schwartzman JD (1997) A novel class of unconventional myosins from *Toxoplasma gondii*. *J Mol Biol.* 271(1):139-146
- Higgins SJ, Xing K, Kim H, Kain DC, Wang F, Dhabangi A, Musoke C, Cserti-Gazdewich CM, Tracey KJ, Kain KC, and Liles WC (2013) Systemic release of high mobility group box 1 (HMGB1) protein is associated with severe and fatal *Plasmodium falciparum* malaria. *Malaria Journal.* 12:105-113.
- Hiller NL, Bhattacharjee S, van Ooij C, Liolios K, Harrison T, Lopez-Estraño C, and Haldar K (2004) A host-targeting signal in virulence proteins reveals a secretome in malarial infection. *Science*. 306:1934-7.

- Hirai M, Arai M, Kawai S and Matsuoka H (2006) PbGCbeta is essential for *Plasmodium* ookinete motility to invade midgut cell and for successful completion of parasite life cycle in mosquitoes, J. Biochem. 140; 747-757.
- History of Malaria, Centers for Disease Control and Prevention, CDC, Atlanta; Internet website: [www.cdc.gov/malaria/about/history/](http://www.cdc.gov/malaria/about/history/)
- Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, Wincker P, Clark AG, Ribeiro JM, Wides R, Salzberg SL, Loftus B, Yandell M, Majoros WH, Rusch DB, Lai Z, Kraft CL, Abril JF, Anthouard V, Arensburger P, Atkinson PW, Baden H, de Berardinis V, Baldwin D, Benes V, Biedler J, Blass C, Bolanos R, Boscus D, Barnstead M, Cai S, Center A, Chaturverdi K, Christophides GK, Chrystal MA, Clamp M, Cravchik A, Curwen V, Dana A, Delcher A, Dew I, Evans CA, Flanigan M, Grundschober-Freimoser A, Friedli L, Gu Z, Guan P, Guigo R, Hillenmeyer ME, Hladun SL, Hogan JR, Hong YS, Hoover J, Jaillon O, Ke Z, Kodira C, Kokoza E, Koutsos A, Letunic I, Levitsky A, Liang Y, Lin JJ, Lobo NF, Lopez JR, Malek JA, McIntosh TC, Meister S, Miller J, Mobarry C, Mongin E, Murphy SD, O'Brochta DA, Pfannkoch C, Qi R, Regier MA, Remington K, Shao H, Sharakhova MV, Sitter CD, Shetty J, Smith TJ, Strong R, Sun J, Thomasova D, Ton LQ, Topalis P, Tu Z, Unger MF, Walenz B, Wang A, Wang J, Wang M, Wang X, Woodford KJ, Wortman JR, Wu M, Yao A, Zdobnov EM, Zhang H, Zhao Q, Zhao S, Zhu SC, Zhimulev I, Coluzzi M, della Torre A, Roth CW, Louis C, Kalush F, Mural RJ, Myers EW, Adams MD, Smith HO, Broder S, Gardner MJ, Fraser CM, Birney E, Bork P, Brey PT, Venter JC, Weissenbach J, Kafatos FC, Collins FH, and Hoffman SL (2002) The Genome Sequence of the Malaria Mosquito *Anopheles gambiae*. Science. 298(5591):129-149.
- Holt RA and Jones SJM (2008) The new paradigm of flow cell sequencing. Genome Res. 18:839-846.
- Horrocks P, Wong E, Russell K and Emes RD (2009) Control of gene expression in *Plasmodium falciparum*--Ten years on. Mol Biochem Parasitol, 164:9-25.
- Hunt P, Afonso A, Creasey A, Culleton R, Sidhu AB, Logan J, Valderramos SG, McNae I, Cheesman S, do Rosario V, Carter R, Fidock DA, and Cravo P (2007) Gene encoding a deubiquitinating enzyme is mutated in artesunate- and chloroquine-resistant rodent malaria parasites. Mol Microbiol. 65(1):27-40.
- Hunt P, Martinelli A, Modrzynska K, Borges S, Creasey A, Rodrigues L, Beraldi D, Loewe L, Fawcett R, Kumar S, Thomson M, Trivedi U, Otto TD, Pain A,

- Blaxter M, and Cravo P (2010) Experimental evolution, genetic analysis and genome re-sequencing reveal the mutation conferring artemisinin resistance in an isogenic lineage of malaria parasites. *BMC Genomics*. 11:499.
- Huse SM, Huber JA, Morrison HG, Sogin ML and Welch DM (2007) Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol*. 8(7):R143.
- Huynh MH, Opitz C, Kwok LY, Tomley FM, Carruthers VB and Soldati D (2004) Trans-genera reconstitution and complementation of an adhesion complex in *Toxoplasma gondii*. *Cell. Microbiol*. 6: 771-782.
- Huynh MH, and Carruthers VB (2006) *Toxoplasma* MIC2 is a major determinant of invasion and virulence. *PLoS Pathog*. 2:e84.
- Ikadai H, Shaw Saliba K, Kanzok SM, McLean KJ, Tanaka TQ, Cao J, Williamson KC, and Jacobs-Lorena M (2013) Transposon mutagenesis identifies genes essential for gametocytogenesis. *Proc Natl Acad Sci U S A*. 110(18):E1676-84.
- Illumina website <http://www.illumina.com>.
- Illumina website:  
[www.illumina.com/technology/mate\\_pair\\_sequencing\\_assay.ilmn](http://www.illumina.com/technology/mate_pair_sequencing_assay.ilmn)
- Illumina, Analyst Day, September 15th 2007, Mandarin Oriental, New York, NY.
- Inselburg J (1983) Stage-specific inhibitory effect of cyclic AMP on asexual maturation and gametocyte formation of *Plasmodium falciparum*. *J Parasitol*. 69:592-597.
- Iwanaga S, Kaneko I, Kato T, and Yuda M (2012) Identification of an AP2-family protein that is critical for malaria liver stage development. *PLoS One*. 7(11):e47557.
- Iyer VR, Horak CE, Scafe CS, Botstein D, Snyder M, and Brown PO (2001) Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* 409: 533-538.
- Iyer LM, Anantharaman V, Wolf MY, and Aravind L (2008) Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. *Int J Parasitol*. 38:1-31.
- Jambou R, El-Assaad F, Combes V, and Grau GE (2011) In vitro culture of *Plasmodium berghei*-ANKA maintains infectivity of mouse erythrocytes inducing cerebral malaria. *Malar J*. 25;10:346.
- Janse CJ, Van Der Klooster PFJ, van Der Kaay HJ, van Der Ploeg M and Overdulver JP (1986) Rapid repeated DNA replication during

- microgametogenesis and DNA synthesis in young zygotes of *Plasmodium berghei*. Trans. Roy. Soc. Trop. Med. Hyg. 80;154-57.
- Janse CJ, Boorsma EG, Ramesar J, Van Vianen PH, Van Der Meer R, Zenobi P, Casaglia O, Mons B and Van Den Berg FM (1989) *Plasmodium berghei*: Gametocyte production, DNA content, and chromosome-size polymorphisms during asexual multiplication in vivo. Experimental Parasitology. 68:274-282.
- Janse CJ and Mons B (1992a) Deletion, insertion and translocation of DNA sequences contribute to chromosome size polymorphism in *Plasmodium berghei*. Mem. Inst. Oswaldo Cruz. 87:Suppl. III;95-100.
- Janse CJ, Ramesar J, Van Den Berg FM and Mons B (1992b) *Plasmodium berghei*: In Vivo Generation and Selection of Karyotype Mutants and Non-Gametocyte Producer Mutants. Experimental Parasitology. 74:1-10.
- Janse CJ, Carlton JMR, Walliker D and Waters AP (1994) Conserved location of genes on polymorphic chromosomes of four species of malaria parasites. Mol. Biochem. Parasitol. 68: 285-296.
- Janse CJ and Waters AP (1995) *Plasmodium berghei*: The Application of Cultivation and Purification Techniques to Molecular Studies of Malaria Parasites. 11(4);138-143.
- Janse CJ and Waters AP (2004) Sexual development of malaria parasites. In Malaria Parasites: Genomes and Molecular Biology. Waters, A.P., and Janse, C.J. (eds). Norwich: Caister Academic Press, pp. 445-474.
- Janse CJ, Ramesar J, and Waters AP (2006) High-efficiency transfection and drug selection of genetically transformed blood stages of the rodent malaria parasite *Plasmodium berghei*. Nat Protoc. 1(1):346-56.
- Janse CJ and Waters AP (2007) The exoneme helps malaria parasites to break out of blood cells. Cell. 131:1036-1038.
- Janse CJ, Kroeze H, van WA, Mededovic S, Fonager J, Franke-Fayard B, Waters AP, and Khan SM (2011) A genotype and phenotype database of genetically modified malaria-parasites. Trends Parasitol. 27: 31-39.
- Jeck WR, Reinhardt JA, Baltrus DA, Hickenbotham MT, Magrini V, Mardis ER, Dangl JL and Jones CD (2007) Extending assembly of short DNA sequences to handle error. Bioinformatics. 23(21): 2942-2944.
- Jiang L, Mu J, Zhang Q, Ni T, Srinivasan P, Rayavara K, Yang W, Turner L, Lavstsen T, Theander TG, Peng W, Wei G, Jing Q, Wakabayashi Y, Bansal A, Luo Y, Ribeiro JMC, Scherf A, Aravind L, Zhu J, Zhao K and Miller LH (2013)

- PfSETvs methylation of histone H3K36 represses virulence genes in *Plasmodium falciparum*. *Nature*. 499: 223-229.
- Johnson DS, Mortazavi A, Myers RM and Wold B (2007) Genome-Wide Mapping of in Vivo Protein-DNA Interactions. *Science*. 316:1497-1502.
- Joseph R, Orlov YL, Huss M, Sun W, Kong SL, Ukil L, Pan YF, Li G, Lim M, Thomsen JS, Ruan Y, Clarke ND, Prabhakar S, Cheung E, and Liu ET (2010) Integrative model of genomic factors for determining binding site selection by estrogen receptor- $\alpha$ . *Mol Syst Biol*. 6: 456.
- Jost JP, Munch O and Andersson T (1991) Study of protein-DNA interactions by surface plasmon resonance (real time kinetics). *Nucleic Acids Res*. 19: 2788.
- Kadekoppala M, Cheresh P, Catron D, Ji DD, Deitsch K, Wellem's TE, Seifert HS, and Halder K (2001) Rapid recombination among transfected plasmids, chimeric episome formation and trans gene expression in *Plasmodium falciparum*. *Mol. Biochem. Parasitol*. 112: 211-218.
- Kadota K, Ishino T, Matsuyama T, Chinzei Y, and Yuda M (2004) Essential role of membrane-attack protein in malarial transmission to mosquito host. *Proc Natl Acad Sci U S A*. 101(46):16310-5.
- Kafsack BFC, Rovira-Graells N, Clark TG, Bancells C, Crowley VM, Campino SG, Williams AE, Drought LG, Kwiatkowski DP, Baker DA, Cortes A and Llinas M (2014) A transcriptional switch underlies commitment to sexual development in malaria parasites. *Nature*. 507(7491):248-52
- Kaplan T, Li XY, Sabo PJ, Thomas S, Stamatoyannopoulos JA, Biggin MD, and Eisen MB (2011) Quantitative models of the mechanisms that control genome-wide patterns of transcription factor binding during early *Drosophila* development. *PLoS Genet* 7: e1001290.
- Kappe SH, Kaiser K, and Matuschewski K (2003) The *Plasmodium* sporozoite journey: a rite of passage. *Trends Parasitol*. 19(3):135-43.
- Karamouzis MV, and Papavassiliou AG (2011) Transcription Factor Networks as Targets for Therapeutic Intervention of Cancer: The Breast Cancer Paradigm. *Mol. Med*. 17 (11-12) 1133-1136.
- Kats LM, Black CG, Proellocks NI and Coppel RL (2006) *Plasmodium* rhoptries: how things went pear-shaped. *Trends Parasitol*. 22(6);269-76.
- Kaushal DC, Carter R, Miller LH and Krishna G (1980) Gametocytogenesis by malaria parasites in continuous culture. *Nature*, 286:490-492.

- Kawamoto E, Alejo-Blanco R, Fleck SL, Kawamoto Y, and Sinden RE (1990) Possible roles of Ca<sup>2+</sup> and cGMP as mediators of the exflagellation of *Plasmodium berghei* and *Plasmodium falciparum*, Mol. Biochem. Parasitol. 42;101-108.
- Khalil AS, Lu TK, Bashor CJ, Ramirez CL, Pyenson NC, Joung JK, and Collins JJ (2012) A synthetic biology framework for programming eukaryotic transcription functions. Cell. 150(3):647-58.
- Khan SM, Franke-Fayard B, Mair GR, Lasonder E, Janse CJ, Mann M, and Waters AP (2005) Proteome analysis of separated male and female gametocytes reveals novel sex-specific *Plasmodium* biology. Cell, 121;675-687.
- Killick-Kendrick R (1974) Parasitic protozoa of the blood of rodents: a revision of *Plasmodium berghei*. Parasitology 69, 225-237.
- Knuepfer E, Suleyman O, Dluzewski AR, Straschil U, O'Keeffe AH, Ogun SA, Green JL, Grainger M, Tewari R, and Holder AA (2013) RON12, a novel *Plasmodium*-specific rhoptry neck protein important for parasite proliferation. Cell Microbiol. 16(5):657-72.
- Kongkasuriyachai D, Fujioka H, and Kumar N (2004) Functional analysis of *Plasmodium falciparum* parasitophorous vacuole membrane protein (Pfs16) during gametocytogenesis and gametogenesis by targeted gene disruption. Mol Biochem Parasitol. 133(2):275-85.
- Kooij TWA, Carlton JM, Bidwell SL, Hall N, Ramesar J, Janse CJ, and Waters AP (2005a) A *Plasmodium* Whole-Genome Synteny Map: Indels and Synteny Breakpoints as Foci for Species-Specific Genes. PLoS Pathog. 1:4;e44.
- Kooij TWA, Franke-Fayard B, Renz J, Kroeze H, van Dooren MW, Ramesar J, Augustijn KD, Janse CJ, and Waters AP (2005b) *Plasmodium berghei* alpha-tubulin II: a role in both male gamete formation and asexual blood stages. Mol Biochem Parasitol. 144:16-26.
- Kooij TWA, Janse CJ, and Waters AP (2006) *Plasmodium* post-genomics: better the bug you know? Nature Reviews Microbiology. 4(5);344-357.
- Kooij TWA and Matuschewski K (2007) Triggers and tricks of *Plasmodium* sexual development. Current Opinion in Microbiology. 10:547-553.
- Kooij TWA, Rauch MM, and Matuschewski K (2012) Expansion of experimental genetics approaches for *Plasmodium berghei* with versatile transfection vectors. Molecular & Biochemical Parasitology. 185:19-26.

- Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, Simons JF, Kim P M, Palejev D, Carriero NJ, Du L, Taillon BE, Chen Z, Tanzer A, Saunders AC, Chi J, Yang F, Carter NP, Hurles ME, Weissman SM, Harkins TT, Gerstein MB, Egholm M, and Snyder M (2007) Paired-end mapping reveals extensive structural variation in the human genome. *Science*. 318:420-426.
- Kornberg RD (1977) Structure of chromatin. *Annu Rev Biochem* 46: 931-954.
- Kornberg RD and Lorch Y (1999) Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell*. 6;98(3):285-94
- Kornberg RD (2007) The molecular basis of eukaryotic transcription. *Proceedings of the National Academy of Sciences of the USA*. 104(32): 12955-12961.
- Korochkina S, Barreau C, Pradel G, Jeffery E, Li J, Natarajan R, Shabanowitz J, Hunt D, Frevert U, and Vernick KD (2006) A mosquito-specific protein family includes candidate receptors for malaria sporozoite invasion of salivary glands. *Cell Microbiol*. 8(1):163-75.
- Kouzarides T (2007) Chromatin modifications and their function. *Cell* 128: 693-705.
- Krause L, Diaz NN, Bartels D, Edwards RA, Pühler A, Rohwer F, Meyer F and Stoye J (2006) Finding novel genes in bacterial communities isolated from the environment. *Bioinformatics*. 22:e281-e289.
- Krizek BA and Sulli C (2006) Mapping sequences required for nuclear localization and the transcriptional activation function of the Arabidopsis protein AINTEGUMENTA. *Planta*. 224:612-21.
- Kumar K, Singal A, Rizvi M, and Chauhan V (2008) High mobility group box (HMGB) proteins of *Plasmodium falciparum*: DNA binding proteins with pro-inflammatory activity. *Parasitology International*, 57(2); 150-157.
- Kurokawa H, Motohashi H, Sueno S, Kimura M, Takagawa H, Kanno Y, Yamamoto M and Tanaka T (2009) Structural basis of alternative dna recognition by maftranscription factors. *Mol Cell Biol*. 29(23):6232-6244.
- LaCount D, Vignali M, Chettier R, Phansalkar A, Bell R, Hesselberth JR, Schoenfeld LW, Ota I, Sahasrabudhe S, Kurschner C, Fields S, and Hughes RE (2005) A protein interaction network of the malaria parasite *Plasmodium falciparum*. *Nature*. 438(7064);103-107.
- Lacroix C, Giovannini D, Combe A, Bargieri DY, Späth S, Panchal D, Tawk L, Thiberge S, Carvalho TG, Barale JC, Bhanot P, and Ménard R (2011) FLP/FRT-



- mediated conditional mutagenesis in pre-erythrocytic stages of *Plasmodium berghei*. Nat Protoc. 6(9):1412-28.
- Lal K, Delves MJ, Bromley E, Wastling JM, Tomley FM, and Sinden RE (2009) *Plasmodium* male development gene-1 (mdv-1) is important for female sexual development and identifies a polarised plasma membrane during zygote development. Int J Parasitol. 39:755-761.
- Langhorne J, Buffet P, Galinski M, Good M, Harty J, Leroy D, Mota MM, Pasini E, Renia L, Riley E, Stins M, and Duffy P (2011) The relevance of non-human primate and rodent malaria models for humans. Malar J. 10(1):23.
- Lavazec C, Moreira CK, Mair GR, Waters AP, Janse CJ, and Templeton TJ (2009) Analysis of mutant *Plasmodium berghei* parasites lacking expression of multiple PbCCp genes. Mol Biochem Parasitol. 163(1):1-7.
- Le Roch KG, Zhou Y, Blair PL, Grainger M, Moch JK, Haynes JD, De La Vega P, Holder AA, Batalov S, Carucci DJ and Winzeler EA (2003) Discovery of gene function by expression profiling of the malaria parasite life cycle. Science. 301: 1503-1508.
- Le Roch KG, Johnson J, Florens L, Zhou Y, Santosyan A, Grainger M, Yan SF, Williamson KC, Holder AA, Carucci DJ, Yates JR, and Winzeler EA (2004) Global analysis of transcript and protein levels across the *Plasmodium falciparum* life cycle. Genome Research. 14(11): 2308-2318.
- Leander BS, Clopton RE and Keeling PJ (2003) Phylogeny of gregarines (Apicomplexa) as inferred from small-subunit rDNA and beta-tubulin. Int. J. Syst. Evol. Microbiol. 53: 345-354.
- Lee MC and Fidock DA (2008) Arresting malaria parasite egress from infected red blood cells. Nat. Chem. Biol. (4):3;161-162.
- Leiden Malaria Research Group website:  
<http://www.lumc.nl/con/1040/81028091348221/810281121192556/811070736212556/>. Leiden University Medical Center. Leiden. The Netherlands.
- Lemon B and Tjian R (2000) Orchestrated response: a symphony of transcription factors for gene control. Genes Dev. 14(20):2551-69.
- Lespinet O, Wolf YI, Koonin EV and Aravind L (2002) The role of lineage-specific gene family expansion in the evolution of eukaryotes. Genome Res. 12: 1048-1059.
- Levine M and Tjian R (2003) Transcription regulation and animal diversity. Nature. 424: 147-151.

- Lew VL (2001) Packaged merozoite release without immediate host cell lysis. *TRENDS in Parasitology*. 17(9);401-403.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, and Durbin R; 1000 Genome Project Data Processing Subgroup. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 25(16):2078-9.
- Li JL, Targett GA and Baker DA (2001) Primary structure and sexual stage-specific expression of a LAMMER protein kinase of *Plasmodium falciparum*. *Int J Parasitol*. 31:387-392.
- Li R, Li Y, Fang X, Yang H, Wang J, Kristiansen K, and Wang J (2009) SNP detection for massively parallel whole-genome resequencing. *Genome Res*. 19(6):1124-32.
- Li XY, MacArthur S, Bourgon R, Nix D, Pollard DA, Iyer VN, Hechmer A, Simirenko L, Stapleton M, Luengo Hendriks CL, Chu HC, Ogawa N, Inwood W, Sementchenko V, Beaton A, Weizmann R, Celniker SE, Knowles DW, Gingeras T, Speed TP, Eisen MB, and Biggin MD (2008) Transcription factors bind thousands of active and inactive regions in the *Drosophila* blastoderm. *PLoS Biol*. 6, e27
- Limviroj W, Yano K, Yuda M, Ando K, and Chinzei Y (2002) Immuno-electron microscopic observation of *Plasmodium berghei* CTRP localization in the midgut of the vector mosquito *Anopheles stephensi*. *J Parasitol*. 88(4):664-72.
- Lin JW, Annoura T, Sajid M, Chevalley-Maurel S, Ramesar J, Klop O, Franke-Fayard BM, Janse CJ, and Khan SM (2011) A novel 'gene insertion/marker out' (GIMO) method for transgene expression and gene complementation in rodent malaria parasites. *PLoS One*. 6(12):e29289.
- Lin S and Riggs AD (1975) The general affinity of lac repressor for *E. coli* DNA: implications for gene regulation in procaryotes and eucaryotes. *Cell*. 4: 107-111.
- Lindner SE, De Silva EK, Keck JL and Llinas M (2010) Structural determinants of DNA binding by a *P. falciparum* ApiAP2 transcriptional regulator. *Journal of Molecular Biology*. 395(3) 558-567.
- Lingnau A, Margos G, Maier WA, and Seitz HM (1993) The effects of hormones on the gametocytogenesis of *Plasmodium falciparum* in vitro. *Appl Parasitol*. 34(3):153-60.

- Liu Y, Tewari R, Ning J, Blagborough AM, Garbom S, Pei J, Grishin NV, Steele RE, Sinden RE, Snell WJ, and Billker O (2008) The conserved plant sterility gene HAP2 functions after attachment of fusogenic membranes in *Chlamydomonas* and *Plasmodium* gametes. *Genes Dev.* 22(8):1051-68.
- Llinas M, Bozdech Z, Wong ED, Adai AT and DeRisi JL (2006) Comparative whole genome transcriptome analysis of three *Plasmodium falciparum* strains. *Nucleic Acids Res.* 34: 1166-1173.
- Llinas M, Deitsch KW and Voss TS (2008) *Plasmodium* gene regulation: far more to factor in. *Trends in parasitology.* 24(12):551-6.
- Lobo CA and Kumar N (1998) Sexual Differentiation and Development in the Malaria Parasite. *Parasitology Today.* 14(4);146-150.
- Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D and Darnell JE (1999) *Molecular Cell Biology.* W.H. Freeman & Co., NY. ISBN-10: 0-7167-3136-3.
- Lopez-Bigas N, De S, and Teichmann SA (2008) Functional protein divergence in the evolution of *Homo sapiens*. *Genome Biol.* 9(2):R33.
- Lopez-Rubio JJ, Gontijo AM, Nunes MC, Issar N, Hernandez Rivas R, and Scherf A (2007) 5' flanking region of var genes nucleate histone modification patterns linked to phenotypic inheritance of virulence traits in malaria parasites. *Mol. Microbiol.* 66: 1296-1305.
- Lopez-Rubio JJ, Mancio-Silva L, and Scherf A (2009) Genome-wide Analysis of Heterochromatin Associates Clonally Variant Gene Regulation with Perinuclear Repressive Centers in Malaria Parasites. *Cell Host and Microbe.* 5;2:179-190.
- Luger K, Ma`der AW, Richmond RK, Sargent DF, and Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389: 251-255.
- MacArthur S, Li XY, Li J, Brown JB, Chu HC, Zeng L, Grondona BP, Hechmer A, Simirenko L, Keränen SV, Knowles DW, Stapleton M, Bickel P, Biggin MD, and Eisen MB (2009) Developmental roles of 21 *Drosophila* transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions. *Genome Biol.* 10, R80
- MacQuarrie KL, Fong AP, Morse RH, and Tapscott SJ (2011) Genome-wide transcription factor binding: beyond direct target regulation. *Trends Genet.* 27(4):141-8.

- Mahairaki V, Voyatzi T, Sidén-Kiamos I, and Louis C (2005) The *Anopheles gambiae* gamma1 laminin directly binds the *Plasmodium berghei* circumsporozoite- and TRAP-related protein (CTRP). *Mol Biochem Parasitol.* 140(1):119-21.
- Maier AG, Braks JA, Waters AP, and Cowman AF (2006) Negative selection using yeast cytosine deaminase/uracil phosphoribosyl transferase in *Plasmodium falciparum* for targeted gene deletion by double crossover recombination. *Mol Biochem Parasitol.* 150(1):118-21
- Mair GR, Braks JAM, Garver LS, Wiegant JCAG, Hall N, Dirks RW, Khan SM, Dimopoulos G, Janse CJ, and Waters AP (2006) Regulation of sexual development of *Plasmodium* by translational repression. *Science.* 313:667-669.
- Mair GR, Lasonder E, Garver LS, Franke-Fayard BMD, Carret CK, Wiegant JC, Dirks RW, Dimopoulos G, Janse CJ, and Waters AP (2010) Universal Features of Post-Transcriptional Gene Regulation Are Critical for *Plasmodium* Zygote Development. *PLoS Pathog.* 6(2):e1000767.
- Mancio-Silva L, Lopez-Rubio JJ, Claes A, and Scherf A (2013) Sir2a regulates rDNA transcription and multiplication rate in the human malaria parasite *Plasmodium falciparum*. *Nat. Commun.* 4:1530.
- Mantel PY, Hoang AN, Goldowitz I, Potashnikova D, Hamza B, Vorobjev I, Ghiran I, Toner M, Irimia D, Ivanov AR, Barteneva N, and Marti M (2013) Malaria-infected erythrocyte-derived microvesicles mediate cellular communication within the parasite population and with the host immune system. *Cell Host Microbe.* 13(5):521-34.
- March S, Ng S, Velmurugan S, Galstian A, Shan J, Logan DJ, Carpenter AE, Thomas D, Sim BK, Mota MM, Hoffman SL, and Bhatia SN (2013) A Microscale Human Liver Platform that Supports the Hepatic Stages of *Plasmodium falciparum* and vivax. *Cell Host Microbe.* 14(1):104-15.
- Mardis ER (2008) The impact of next-generation sequencing technology on genetics. *Trends Genet.* 24:133-41.
- Marti M, Good RT, Rug M, Knuepfer E and Cowman AF (2004) Targeting malaria virulence and remodeling proteins to the host erythrocyte. *Science.* 306:1930-3
- Martinelli A, Henriques G, Cravo P, and Hunt P (2011) Whole genome re-sequencing identifies a mutation in an ABC transporter (mdr2) in a

- Plasmodium chabaudi* clone with altered susceptibility to antifolate drugs. Int J Parasitol. 41(2):165-71.
- Masumoto H, Hawke D, Kobayashi R, and Verreault A (2005) A role for cell-cycle-regulated histone H3 lysine 56 acetylation in the DNA damage response. Nature. 436(7048): 294-8.
- Maswoswe SM, Peters W, and Warhurst DC (1985) Corticosteroid stimulation of the growth of *Plasmodium falciparum* gametocytes in vitro. Ann Trop Med Parasitol. 79(6):607-16.
- Matsuoka H, Yoshida S, Hirai M, and Ishii A (2002) A rodent malaria, *Plasmodium berghei*, is experimentally transmitted to mice by merely probing of infective mosquito, *Anopheles stephensi*. Parasitol Int 51:17-23.
- Matuschewski K, and Schuler H (2008) Actin/myosin-based gliding motility in apicomplexan parasites. Subcell Biochem. 47:110-120.
- Matz JM, Matuschewski K and Kooij TWA (2013) Two putative protein export regulators promote malaria blood stage development *in vivo*. Mol Biochem Parasitol. 191(1):44-52
- McNally J, O'Donovan SM, and Dalton JP (1992) *Plasmodium berghei* and *Plasmodium chabaudi chabaudi*: development of simple *in vitro* erythrocyte invasion assays. Parasitology. 105 (Pt 3):355-62.
- McRobert L, Taylor CJ, Deng W, Fivelman QL, Cummings RM, Polley SD, Billker O, and Baker DA (2008) Gametogenesis in malaria parasites is mediated by the cGMP-dependent protein kinase. PLoS Biol 6(6): e139.
- Medica DL, and Sinnis P (2005) Quantitative dynamics of *Plasmodium yoelii* sporozoite transmission by infected anopheline mosquitoes feeding on vertebrate hosts. Infect Immun. 73:4363-4369.
- Meis JF, Pool G, van Gemert GJ, Lensen AH, Ponnudurai T, and Meuwissen JH (1989) *Plasmodium falciparum* ookinetes migrate intercellularly through *Anopheles stephensi* midgut epithelium. Parasitol Res. 76(1):13-9.
- Meis JF, Wismans PG, Jap PH, Lensen AH, and Ponnudurai T (1992) A scanning electron microscopic study of the sporogonic development of *Plasmodium falciparum* in *Anopheles stephensi*. Acta Trop. 50(3):227-36.
- Meissner M, Schliiter D, and Soldati D (2002) Role of *Toxoplasma gondii* myosin A in powering parasite gliding and host cell invasion. Science. 298(5594):837-840.

- Meissner M, Krejany E, Gilson PR, de Koning-Ward TF, Soldati D, and Crabb BS (2005) Tetracycline analogue-regulated transgene expression in *Plasmodium falciparum* blood stages using *Toxoplasma gondii* transactivators. *Proc Natl Acad Sci USA*. 102:2980-5.
- Meissner M, Ferguson DJP, and Frischknecht F (2013) Invasion factors of apicomplexan parasites: essential or redundant? *Current Opinion in Microbiology*. 16:438-444.
- Ménard R (2005) Knockout malaria vaccine? *Nature*. 433:113-114.
- Mendis K, Rietveld A, Warsame M, Bosman A, Greenwood B and Wernsdorfer WH (2009) From malaria control to eradication: The WHO perspective. *Tropical Medicine & International Health*. 14:7;802-809.
- Menegon M, Sannella AR, Majori G, and Severini C (2008) Detection of novel point mutations in the *Plasmodium falciparum* ATPase6 candidate gene for resistance to artemisinins. *Parasitol Int*. 57(2):233-5.
- Mercier C, Adjogble KD, Däubener W and Delauw MF (2005) Dense granules: are they key organelles to help understand the parasitophorous vacuole of all apicomplexa parasites? *Int J Parasitol*. 35(8);829-49.
- Metzker ML (2010) Sequencing technologies - the next generation. *Nature Reviews Genetics*. 11:31-46.
- Miao J, Fan Q, Cui L, Li J, Li J, and Cui L (2006) The malaria parasite *Plasmodium falciparum* histones: organization, expression, and acetylation. *Gene*. 369: 53-65.
- Miao J, Li J, Fan Q, Li X, Li X, and Cui L (2010) The Puf-family RNA-binding protein PfPuf2 regulates sexual development and sex differentiation in the malaria parasite *Plasmodium falciparum*. *J Cell Sci*. 123: 1039-1049.
- Micks DW (1947) A loss of gametocytes in *Plasmodium elongatum*. *J Parasitol*. 33(6):499-505.
- Mideo N, Kennedy DA, Carlton JM, Bailey JA, Juliano JJ, and Read AF (2013) Ahead of the curve: next generation estimators of drug resistance in malaria infections. *Trends Parasitol*. 29(7):321-8.
- Mikolajczak SA, Silva-Rivera H, Peng X, Tarun AS, Camargo N, Jacobs-Lorena V, Daly TM, Bergman LW, de la Vega P, Williams J, Aly AS, and Kappe SH (2008) Distinct malaria parasite sporozoites reveal transcriptional changes that cause differential tissue infection competence in the mosquito vector and mammalian host. *Mol Cell Biol*. 28:6196-207.

- Miller LH, and Su X (2011) Artemisinin: discovery from the Chinese herbal garden. *Cell*. 146(6):855-8.
- Miller SK, Good RT, Drew DR, Delorenzi M, Sanders PR, Hodder AN, Speed TP, Cowman AF, de Koning-Ward TF and Crabb BS (2002) A subset of *Plasmodium falciparum* SERA genes are expressed and appear to play an important role in the erythrocytic cycle. *J Biol Chem*. 277(49):47524-32.
- Miotto O, Almagro-Garcia J, Manske M, Macinnis B, Campino S, Rockett KA, Amaratunga C, Lim P, Suon S, Sreng S, Anderson JM, Duong S, Nguon C, Chuor CM, Saunders D, Se Y, Lon C, Fukuda MM, Amenga-Etego L, Hodgson AV, Asoala V, Imwong M, Takala-Harrison S, Nosten F, Su XZ, Ringwald P, Arie F, Dolecek C, Hien TT, Boni MF, Thai CQ, Amambua-Ngwa A, Conway DJ, Djimdé AA, Doumbo OK, Zongo I, Ouedraogo JB, Alcock D, Drury E, Auburn S, Koch O, Sanders M, Hubbard C, Maslen G, Ruano-Rubio V, Jyothi D, Miles A, O'Brien J, Gamble C, Oyola SO, Rayner JC, Newbold CI, Berriman M, Spencer CC, McVean G, Day NP, White NJ, Bethell D, Dondorp AM, Plowe CV, Fairhurst RM, and Kwiatkowski DP (2013) Multiple populations of artemisinin-resistant *Plasmodium falciparum* in Cambodia. *Nat Genet*. 45(6):648-55.
- Mitchell GH, Thomas AW, Margos G, Dluzewski AR, and Bannister LH (2004) Apical membrane antigen 1, a major malaria vaccine candidate, mediates the close attachment of invasive merozoites to host red blood cells. *Infect Immun*. 72(1):154-8.
- Modrzynska K, Creasey A, Loewe L, Cezard T, Trindade Borges S, Martinelli A, Rodrigues L, Cravo P, Blaxter M, Carter R, and Hunt P (2012) Quantitative genome re-sequencing defines multiple mutations conferring chloroquine resistance in rodent malaria. *BMC Genomics*. 13:106.
- Molina-Cruz A, Garver LS, Alabaster A, Bangiolo L, Haile A, Winikor J, Ortega C, van Schaijk BCL, Sauerwein RW, Taylor-Salmon E, and Barillas-Mur C (2013) The Human Malaria Parasite Pfs47 Gene Mediates Evasion of the Mosquito Immune System. *Science*. 340(6135) 984-987.
- Molloy S (2010) The *Plasmodium* stage manager. *Nat Rev Microbiol*. 8(9):611.
- Mons B, Janse CJ, Boorsma EG and Van Der Kaay HJ (1985) Synchronized erythrocytic schizogony and gametocytogenesis of *Plasmodium berghei* in vivo and in vitro. *Parasitology*. 91:423-430.
- Mons, B (1986) Intraerythrocytic differentiation of *Plasmodium berghei*. *Acta Leidensia*. 54:1-83.

- Moon RW, Taylor CJ, Bex C, Schepers R, Goulding D, Janse CJ, Waters AP, Baker DA and Billker O (2009) A cyclic GMP signalling module that regulates gliding motility in malaria parasite, *PLoS Pathog.* 5 e1000599.
- Moore MJ, Dhingra A, Soltis PS, Shaw R, Farmerie WG, Foltz KM and Soltis DE (2006) Rapid and accurate pyrosequencing of angiosperm plastid genomes. *BMC Plant Biol.* 6:17.
- Moorthy VS, Newman RD, Duclos P, Okwo-Bele JM, and Smith PG (2013a) Assessment of the RTS,S/AS01 malaria vaccine. *Lancet Infect Dis.* 13(4):280-2.
- Moorthy VS, Newman RD, and Okwo-Bele JM (2013b) Malaria vaccine technology roadmap. *Lancet* 382(9906):1700-1701.
- Moraes CB, Dorval T, Contreras-Dominguez M, Dossin FdM, Hansen MAE, Genovesio A, and Freitas-Junior LH (2013) Transcription Sites Are Developmentally Regulated during the Asexual Cycle of *Plasmodium falciparum*. *PLoS ONE* 8(2): e55539.
- Morahan BJ, Strobel C, Hasan U, Czesny B, Mantel PY, Marti M, Eksi S, and Williamson KC (2011) Functional Analysis of the Exported Type IV HSP40 Protein PfGECO in *Plasmodium falciparum* Gametocytes. *Eukaryot Cell* 10: 1492-1503.
- Morrisette NS, and Sibley LD (2002) Cytoskeleton of apicomplexan parasites. *Microbiol Mol Biol Rev*; 66(l):21-38.
- Mota MM, Pradel G, Vanderberg JP, Hafalla JCR, Frevert U, Nussenzweig RS, Nussenzweig V, and Rodriguez A (2001) Migration of *Plasmodium* sporozoites through cells before infection. *Science.* 291;141-144.
- Mota MM, Thathy V, Nussenzweig RS, and Nussenzweig V (2001) Gene targeting in the rodent malaria parasite *Plasmodium yoelii*. *Mol. Biochem. Parasitol.* 113: 271-278.
- Motard A, Marussig M, Renia L, Baccam D, Landau I, Mattei D, Targett G and Mazier D (1995) Immunization with the malaria heat shock like protein hsp70-1 enhances transmission to the mosquito. *Int Immunol*, 7:147-150.
- Mourier T, Carret C, Kyes S, Christodoulou Z, Gardner PP, Jeffares DC, Pinches R, Barrell B, Berriman M, Griffiths-Jones S, Ivens A, Newbold C, and Pain A (2008) Genome-wide discovery and verification of novel structured RNAs in *Plasmodium falciparum*. *Genome Res.* 18: 281-292.



- Mueller AK, Kohlhepp F, Hammerschmidt C, and Michel K (2010) Invasion of mosquito salivary glands by malaria parasites: prerequisites and defense strategies. *Int J Parasitol.* 40(11):1229-35.
- Muhia DK, Swales CA, Deng W, Kelly JM, and Baker DA (2001) The gametocyte-activating factor xanthurenic acid stimulates an increase in membrane-associated guanylyl cyclase activity in the human malaria parasite *Plasmodium falciparum*, *Mol. Microbiol.* 42: 553e560.
- Muhia DK, Swales CA, Eckstein-Ludwig U, Saran S, Polley SD, Kelly JM, Schaap P, Krishna S and Baker DA (2003) Multiple splice variants encode a novel adenylyl cyclase of possible plastid origin expressed in the sexual stage of the malaria parasite *Plasmodium falciparum*. *J Biol Chem.* 278:22014-22022.
- Muller D, and Stelling J (2009) Precise Regulation of Gene Expression Dynamics Favors Complex Promoter Architectures. *PLoS Comput Biol* 5(1): e1000279.
- Muller K, Matuschewski K, and Silvie O (2011) The Puf-Family RNA-Binding Protein Puf2 Controls Sporozoite Conversion to Liver Stages in the Malaria Parasite. *PLoS ONE* 6(5): e19860.
- Murray CJ, Rosenfeld LC, Lim SS, Andrews KG, Foreman KJ, Haring D, Fullman N, Naghavi M, Lozano R, and Lopez AD (2012) Global malaria mortality between 1980 and 2010: a systematic analysis. *Lancet.* 4:379(9814):413-31.
- Narayan V, Eckert M, Zylicz A, Zylicz M, and Ball KL (2009) Cooperative regulation of the interferon regulatory factor-1 tumor suppressor protein by core components of the molecular chaperone machinery. *J Biol Chem.* 284(38):25889-99.
- Navadgi VM, Chandra BR, Mishra PC, and Sharma A (2006) The two *Plasmodium falciparum* nucleosome assembly proteins play distinct roles in histone transport and chromatin assembly. *J. Biol. Chem.* 281: 16978-16984.
- Neafsey DE (2013) Genome sequencing sheds light on emerging drug resistance in malaria parasites. *Nat Genet.* 45(6):589-90.
- Nelson HC, Finch JT, Luisi BF and Klug A (1987) The structure of an oligo(dA).oligo(dT) tract and its biological implications. *Nature.* 330: 221-226.
- Ngwa CJ, Scheuermayer M, Mair GR, Kern S, Brügl T, Wirth CC, Aminake MN, Wiesner J, Fischer R, Vilcinskas A and Pradel G (2013) Changes in the transcriptome of the malaria parasite *Plasmodium falciparum* during the

- initial phase of transmission from the human to the mosquito. *BMC Genomics*. 14:256
- Nickerson DA, Tobe VO and Taylor SL (1997) PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res.* 25:2745-2751.
- Nijhout MM and Carter R (1978) Gamete development in malaria parasites: bicarbonate-dependent stimulation by pH in vitro. *Parasitology*. 76;39-53.
- O'Donnell R, Preiser PR, Williamson DH, Moore PW, Cowman AF, and Crabb BS (2001) An alteration in concatameric structure is associated with efficient segregation of plasmids in transfected *Plasmodium falciparum* parasites. *Nucleic Acids Res.* 29: 716-724.
- O'Donnell RA, Freitas-Junior LH, Preiser PR, Williamson DH, Duraisingh M, McElwain TF, Scherf A, Cowman AF, and Crabb BS (2002) A genetic screen for improved plasmid segregation reveals a role for Rep20 in the interaction of *Plasmodium falciparum* chromosomes. *Embo J.* 21: 1231-1239.
- Oguariri RM, Dunn JM and Golightly LM (2006) 3' Gene regulatory elements required for expression of the *Plasmodium falciparum* developmental protein, Pfs25. *Mol. Biochem. Parasitol.* 146: 163-172.
- Okulate MA, Kalume DE, Reddy R, Kristiansen T, Bhattacharyya M, Chaerkady R, Pandey A, and Kumar N (2007) Identification and molecular characterization of a novel protein Saglin as a target of monoclonal antibodies affecting salivary gland infectivity of *Plasmodium* sporozoites. *Insect Mol Biol.* 16(6):711-22.
- Olivieri A, Camarda G, Bertuccini L, van de Vegte-Bolmer M, Luty AJ, Sauerwein R, and Alano P (2009) The *Plasmodium falciparum* protein Pfg27 is dispensable for gametocyte and gamete production, but contributes to cell integrity during gametocytogenesis. *Mol. Microbiol.* 73: 180-193.
- Ono T, Nakai T, and Nakabayashi T (1986) Induction of gametocytogenesis in *Plasmodium falciparum* by the culture supernatant of hybridoma cells producing anti-P. *falciparum* antibody. *Biken J.* 29:7781-7786.
- Ono T and Nakabayashi T (1990) Gametocytogenesis induction by ammonium compounds in cultured *Plasmodium falciparum*. *Int J Parasitol.* 20(5):615-8.
- Ono T, Ohnishi Y, Nagamune K, and Kano M (1993) Gametocytogenesis induction by Berenil in cultured *Plasmodium falciparum*. *Exp Parasitol.* 77(1):74-8.

- Orr RY, Philip N, and Waters AP (2012) Improved negative selection protocol for *Plasmodium berghei* in the rodent malarial model. *Malar J.* 11:103.
- Otiene, C (2013) for the RTS,S Clinical Trials Partnership. Efficacy of RTS, S/AS01 Vaccine Candidate against malaria in African infants and children 18 months post-primary vaccination series: a phase III randomized double blind controlled trial (abstract). Multinational Initiative on Malaria, 6th Pan African Malaria Conference, Durban, South Africa, October 6-11, 2013.
- Otto TD, Sanders M, Berriman M, and Newbold C (2010) Iterative Correction of Reference Nucleotides (iCORN) using second generation sequencing technology. *Bioinformatics.* 26(14):1704-7.
- Otto TD, Wilinski D, Assefa S, Keane TM, Sarry LR, Böhme U, Lemieux J, Barrell B, Pain A, Berriman M, Newbold C, and Llinás M (2010) New insights into the blood-stage transcriptome of *Plasmodium falciparum* using RNA-Seq. *Mol Microbiol.* 76: 12-24.
- Otto TD, Dillon GP, Degraeve WS, and Berriman M (2011) RATT: Rapid Annotation Transfer Tool. *Nucleic Acids Res.* 39(9):e57.
- Pace T, Ponzi M, Dore E and Frontali C (1987) Telomeric motifs are present in a highly repetitive element in the *Plasmodium berghei* genome. *Mol. Biochem. Parasitol.* 24:193-202.
- Pace T, Ponzi M, Scotti R and Frontali C (1995) Structure and superstructure of *Plasmodium falciparum* subtelomeric regions. *Molecular and Biochemical Parasitology.* 69:257-268.
- Pace T, Olivieri A, Sanchez M, Albanesi V, Picci L, Siden Kiamos I, Janse CJ, Waters AP, Pizzi E, and Ponzi M (2006) Set regulation in asexual and sexual *Plasmodium* parasites reveals a novel mechanism of stage-specific expression. *Mol Microbiol.* 60: 870-882.
- Pain A, Crossman L and Parkhill J (2005) Comparative Apicomplexan genomics. *Nature Reviews Microbiology.* 3:454-455.
- Pain A and Hertz-Fowler C (2009) *Plasmodium* genomics: latest milestone. *Nature Reviews Microbiology.* 7:180-181.
- Painter HJ, Campbell TL and Llinas M (2011) The Apicomplexan AP2 family: integral factors regulating *Plasmodium* development. *Molecular and Biochemical Parasitology.* 176: 1-7.
- Painter HJ, and Llinás M (2012) Comment on "Emerging functions of transcription factors in malaria parasite". *J Biomed Biotechnol.* 2012:754054.

- Pasvol G, Weatherall DJ, and Wilson RJ (1980) The increased susceptibility of young red cells to invasion by the malarial parasite *Plasmodium falciparum*. Br J Haematol. 45(2):285-95.
- Paton MG, Barker GC, Matsuoka H, Ramesar J, Janse CJ, Waters AP and Sinden RE (1993) Structure and expression of a post-transcriptionally regulated malaria gene encoding a surface protein from the sexual stages of *Plasmodium berghei*. Mol Biochem Parasitol. 59(2):263-75
- Paul RE, Diallo M, and Brey PT (2004) Mosquitoes and transmission of malaria parasites - not just vectors. Malar J. 8(3):39.
- Pedraza JM, and van Oudenaarden A (2005) Noise propagation in gene networks. Science. 307: 1965-1969.
- Perez-Toledo K, Rojas-Meza AP, Mancio-Silva L, Hernandez-Cuevas NA, Delgadillo DM, Vargas M, Martínez-Calvillo S, Scherf A, and Hernandez-Rivas R (2009) *Plasmodium falciparum* heterochromatin protein 1 binds to tri-methylated histone 3 lysine 9 and is linked to mutually exclusive expression of var genes. Nucleic Acids Res. 37: 2596-2606.
- Petter M, Lee CC, Byrne TJ, Boysen KE, Volz J, Ralph SA, Cowman AF, Brown GV, and Duffy MF (2011) Expression of *P. falciparum* var Genes involves exchange of the histone variant H2A.Z at the promoter. PLoS Pathog 7: e1001292.
- Pettersson E, Lundeberg J, and Ahmadian A (2009) Generations of sequencing technologies. Genomics. 93:105-111.
- Pfander C, Anar B, Schwach F, Otto TD, Brochet M, Volkmann K, Quail MA, Pain A, Rosen B, Skarnes W, Rayner JC, and Billker O (2011) A scalable pipeline for highly effective genetic modification of a malaria parasite. Nat Methods. 8(12):1078-82.
- Phillips T, and Hoopes L (2008) Transcription factors and transcriptional control in eukaryotic cells. Nature Education 1(1).
- Pimenta PF, Touray M and Miller LH (1994) The journey of malaria sporozoites in the mosquito salivary gland. J Euk Microbiol. 41: 608-624.
- Pino P, Sebastian S, Kim EA, Bush E, Brochet M, Volkmann K, Kozłowski E, Llinás M, Billker O, and Soldati-Favre D (2012) A Tetracycline-Repressible Transactivator System to Study Essential Genes in Malaria Parasites. Cell Host Microbe. 12(6): 824-834.

- Poespoprodjo JR, Fobia W, Kenangalem E, Lampah DA, Hasanuddin A, Warikar N, Sugiarto P, Tjitra E, Anstey NM and Price RN (2009) Vivax malaria: a major cause of morbidity in early infancy. *Clin. Infect. Dis.* 48: 1704-1712.
- Polson HE and Blackman MJ (2005) A role for poly(dA)poly(dT) tracts in directing activity of the *Plasmodium falciparum* calmodulin gene promoter. *Mol. Biochem. Parasitol.* 141: 179-189.
- Ponnudurai T, Meuwissen JH, Leeuwenberg AD, Verhave JP and Lensen AH (1982) The production of mature gametocytes of *Plasmodium falciparum* in continuous cultures of different isolates infective to mosquitoes. *Trans R Soc Trop Med Hyg.* 76:242-50.
- Ponzi M, Janse CJ, Dore E, Scotti R, Pace T, Reterink TJ, van der Berg FFM and Mons B (1990) Generation of chromosome size polymorphism during *in vivo* mitotic multiplication of *Plasmodium berghei* involves both loss and addition of subtelomeric repeat sequences. *Mol. Biochem. Parasitol.* 41:73-82.
- Ponzi M, Sidén-Kiamos I, Bertuccini L, Currà C, Kroeze H, Camarda G, Pace T, Franke-Fayard B, Laurentino EC, Louis C, Waters AP, Janse CJ, and Alano P (2009) Egress of *Plasmodium berghei* gametes from their host erythrocyte is mediated by the MDV-1/PEG3 protein. *Cell Microbiol.* 11(8):1272-88.
- Pop M and Salzberg SL (2008) Bioinformatics challenges of new sequencing technology. *Trends Genet.* 24(3):142-149.
- Pop M (2009) Genome assembly reborn: recent computational challenges. *Brief Bioinform.* 10;4:354-66.
- Price R, Nosten F, Simpson JA, Luxemburger C, Phaipun L, ter Kuile F, van Vugt M, Chongsuphajaisiddhi T and White NJ (1999) Risk factors for gametocyte carriage in uncomplicated *falciparum* malaria. *Am J Trop Med Hyg.* 60:1019-1023.
- Price RN, Tjitra E, Guerra CA, Yeung S, White NJ, and Anstey NM (2007) Vivax malaria: neglected and not benign. *Am. J. Trop. Med. Hyg.* 77 (Suppl. 6) 79-87.
- Prudêncio M, Mota MM, and Mendes AM (2011) A toolbox to study liver stage malaria. *Trends Parasitol.* 12:565-74.
- Ptashne M (1986) Gene regulation by proteins acting nearby and at a distance. *Nature.* 322: 697-701.
- Ptashne M (1988) How eukaryotic transcriptional activators work. *Nature.* 335: 683-689.

- Ptashne M and Gann A (1997) Transcriptional activation by recruitment. *Nature* 386(6625):569-77.
- Ptashne M and Gann A (2002) *Genes & Signals*, Cold Spring Harbor Laboratory Press. ISBN 978-0-87969-633-7.
- Ptashne M (2007) On the use of the word 'epigenetic'. *Curr Biol* 17: R233-236.
- Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, and Gu Y (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*. 24;13:341.
- Quenault T, Lithgow T, and Traven A (2011) PUF proteins: repression, activation and mRNA localization. *Trends Cell Biol.* 21(2):104-12.
- Raabe AC, Billker O, Vial HJ and Wengelnik K (2009) Quantitative assessment of DNA replication to monitor microgametogenesis in *Plasmodium berghei*. *Mol. Biochem. Parasitol.* 168:172-176.
- Ragoczy T, Bender MA, Telling A, Byron R, and Groudine M (2006) The locus control region is required for association of the murine beta-globin locus with engaged transcription factories during erythroid maturation. *Genes Dev.* 20(11):1447-57.
- Ralph SA, Scheidig-Benatar C, and Scherf A (2005) Antigenic variation in *Plasmodium falciparum* is associated with movement of var loci between subnuclear locations. *Proc. Natl. Acad. Sci. U. S. A.* 102: 5414-5419.
- Ravin NV and Ravin VK (1999) Use of a linear multicopy vector based on the mini-replicon of temperate coliphage N15 for cloning DNA with abnormal secondary structures. *Nucleic Acids Res.* 27, e13.
- Ravin NV, Kuprianov VV, Gilcrease EB and Casjens SR (2003) Bidirectional replication from an internal ori site of the linear N15 plasmid prophage. *Nucleic Acids Res.* 31: 6552-6560.
- Read AF and Day KP (1992) The genetic structure of malaria parasite populations. *Parasitol Today.* 8(7):239-42.
- Rebeiz M, Castro B, Liu F, Yue F, and Posakony JW (2012) Ancestral and conserved cis-regulatory architectures in developmental control genes. *Dev Biol.* 362(2):282-94.
- Rebeiza M, Jikomesa N, Kassner V and Carrolla S (2011) Evolutionary origin of a novel gene expression pattern through co-option of the latent activities of existing regulatory sequences. *PNAS.* 108(25):10036-43.

- Reece SE, Drew DR and Gardner A (2008) Sex ratio adjustment and kin discrimination in malaria parasites. *Nature*. 53: 609-613.
- Reed SG, Bertholet S, Coler RN, and Friede M (2009) New horizons in adjuvants for vaccine development. *Trends Immunol*. 30:23-32.
- Regev-Rudzki N, Wilson DW, Carvalho TG, Sisquella X, Coleman BM, Rug M, Bursac D, Angrisano F, Gee M, Hill AF, Baum J, and Cowman AF (2013) Cell-cell communication between malaria-infected red blood cells via exosome-like vesicles. *Cell*. 153(5):1120-33.
- Reilly HB, Wang H, Steuter JA, Marx AM and Ferdig MT (2007) Quantitative dissection of clone-specific growth rates in cultured malaria parasites. *Int J Parasitol*. 37(14):1599-1607.
- Reininger L, Billker O, Tewari R, Mukhopadhyay A, Fennell C, Dorin-Semblat D, Doerig C, Goldring D, Harmse L, Ranford-Cartwright L, Packer J, and Doerig C (2005) A NIMA-related protein kinase is essential for completion of the sexual cycle of malaria parasites. *J Biol Chem*. 280(36):31957-64.
- Reininger L, Garcia M, Tomlins A, Müller S, and Doerig C (2012) The *Plasmodium falciparum*, Nima-related kinase Pfnek-4: a marker for asexual parasites committed to sexual differentiation. *Malar J*. 11:250.
- Rich SM, and Ayala FJ (2003) Phylogenetics in malaria research: the case for phylogenetics. *Adv Parasitol*. 54:255-280.
- Riechmann JL, and Meyerowitz EM (1998) The AP2/EREBP family of plant transcription factors. *Biol Chem*. 379(6):633-46.
- Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim M, Broun P, Zhang JZ, Ghandehari D, Sherman BK and Yu G (2000) Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *Science*. 290(5499):2105-2110.
- Riglar DT, Richard D, Wilson DW, Boyle MJ, Dekiwadia C, Turnbull L, Angrisano F, Marapana DS, Rogers KL, Whitchurch CB, Beeson JG, Cowman AF, Ralph SA, and Baum J (2011) Super-resolution dissection of coordinated events during malaria parasite invasion of the human erythrocyte. *Cell Host Microbe* 9:9-20.
- Robert V, Read AF, Essong J, Tchuinkam T, Mulder B, Verhave JP, and Carnevale P (1996) Effect of gametocyte sex ratio on infectivity of *Plasmodium falciparum* to *Anopheles gambiae*. *Trans R Soc Trop Med Hyg*. 90(6):621-4.

- Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, Zeng T, Euskirchen G, Bernier B, Varhol R, Delaney A, Thiessen N, Griffith OL, He A, Marra M, Snyder M, and Jones S (2007) Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Methods*. 4: 651-657.
- Rosenberg R, Wirtz RA, Schneider I, and Burge R (1990) An estimation of the number of malaria sporozoites ejected by a feeding mosquito. *Trans Roy Soc Trop Med Hyg.* 84:209-212.
- Rosenfeld N, Young JW, Alon U, Swain PS and Elowitz MB (2005) Gene regulation at the single-cell level. *Science*. 307: 1962-1965.
- Rosenthal PJ (2013) The interplay between drug resistance and fitness in malaria parasites. *Mol Microbiol.* 89(6):1025-38.
- Ruecker A, Shea M, Hackett F, Suarez C, Hirst EM, Milutinovic K, Withers-Martinez C, and Blackman MJ (2012) Proteolytic activation of the essential parasitophorous vacuole cysteine protease ERA6 accompanies malaria parasite egress from its host erythrocyte. *J Biol Chem* 287:37949-63.
- Sakamoto H, Thiberge S, Akerman S, Janse CJ, Carvalho TG, and Menard R (2005) Towards systematic identification of *Plasmodium* essential genes by transposon shuttle mutagenesis. *Nucleic Acids Res.* 33: e174.
- Salcedo-Amaya AM, van Driel MA, Alako BT, Trelle MB, van den Elzen AMG, Cohen AM, Janssen-Megens EM, van de Vegte-Bolmer M, Selzer RR, Iniguez AL, Green RD, Sauerwein RW, Jensen ON and Stunnenberg HG (2009) Dynamic histone H3 epigenome marking during the intraerythrocytic cycle of *Plasmodium falciparum*. *PNAS*. 106;24:9655-9660.
- Sanders PR, Gilson PR, Cantin GT, Greenbaum DC, Nebl T, Carucci DJ, McConville MJ, Schofield L, Hodder AN, Yates JR, and Crabb BS (2005) Distinct protein classes including novel merozoite surface antigens in Raft-like membranes of *Plasmodium falciparum*. *J Biol Chem.* 280(48):40169-76.
- Sauerwein RW, Roestenberg M and Moorthy VS (2011) Experimental human challenge infections can accelerate clinical malaria vaccine development. *Nature Reviews Immunology.* 11: 57-64.
- Schalch T, Duda S, Sargent DF, and Richmond TJ (2005) X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature* 436: 138-141.



- Scherer SW, Lee C, Birney E, Altshuler DM, Eichler EE, Carter NP, Hurles ME and Feuk L (2007) Challenges and standards in integrating surveys of structural variation *Nat Genet.* 39(7) Suppl: S7-15.
- Scherf A, Lopez-Rubio J-J and Riviere L (2008) Antigenic variation in *Plasmodium falciparum*. *Annu. Rev. Microbiol.* 62:445-470.
- Schneider P, Bousema T, Omar S, Gouagna L, Sawa P, Schallig H and Sauerwein R (2006) (Sub)microscopic *Plasmodium falciparum* gametocytaemia in Kenyan children after treatment with sulphadoxine-pyrimethamine monotherapy or in combination with artesunate. *Int. J. Parasitol.* 36:403-408.
- Schneweis S, Maier WA and Seitz HM (1991) Haemolysis of infected erythrocytes - a trigger for formation of *Plasmodium falciparum* gametocytes? *Parasitol Res.* 77:458-460.
- Schoenfelder S, Sexton T, Chakalova L, Cope NF, Horton A, Andrews S, Kurukuti S, Mitchell JA, Umlauf D, Dimitrova DS, Eski CH, Luo Y, Wei CL, Ruan Y, Bieker JJ, and Fraser P (2010) Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat Genet.* 42(1):53-61.
- Scholz SM, Simon N, Lavazec C, Dude MA, Templeton TJ and Pradel G (2008) PfCCp proteins of *Plasmodium falciparum*: gametocyte-specific expression and role in complement-mediated inhibition of exflagellation. *Int J Parasitol.* 38:327-40.
- Schones DE, Cui K, Cuddapah S, Roh TY, Barski A, Wang Z, Wei G and Zhao K (2008) Dynamic regulation of nucleosome positioning in the human genome. *Cell.* 132(5):887-898.
- Sebastian S, Brochet M, Collins MO, Schwach F, Jones ML, Goulding D, Rayner JC, Choudhary JS and Billker O (2012) *Cell Host & Microbe.* 12: 9-19.
- Setty Y, Mayo AE, Surette MG, and Alon U (2003) Detailed map of a cis-regulatory input function. *Proc Natl Acad Sci U S A.* 100: 7702-7707.
- Shaw PJ, Ponmee N, Karoonuthaisiri N, Kamchonwongpaisan S and Yuthavong Y (2007) Characterization of human malaria parasite *Plasmodium falciparum* eIF4E homologue and mRNA 5' cap status. *Mol. Biochem. Parasitol.* 155: 146-155.
- Sherrer RL (2012) Combating malaria: the devil is in the molecular details. *Trends Parasitol.* 28(11):455-6.

- Shirley M, Biggs B, Forsyth K, Brown H, Thompson J, Brown G and Kemp D (1990) Chromosome 9 from independent clones and isolates of *Plasmodium falciparum* undergoes subtelomeric deletions with similar breakpoints *in vitro*. *Molecular and Biochemical Parasitology*. 40: 137-146.
- Shock JL, Fischer KF, and DeRisi JL (2007) Whole-genome analysis of mRNA decay in *Plasmodium falciparum* reveals a global lengthening of mRNA half-life during the intra-erythrocytic development cycle. *Genome Biol.* 8:R134.
- Shulman MJ, Nissen L, and Collins C (1990) Homologous recombination in hybridoma cells: dependence on time and fragment length. *Mol Cell Biol.* 10(9):4466-72.
- Shuman S (2001) Structure, mechanism, and evolution of the mRNA capping apparatus. *Prog. Nucleic Acid Res. Mol. Biol.* 66: 1-40.
- Shuman S (2002) What messenger RNA capping tells us about eukaryotic evolution. *Nat. Rev. Mol. Cell Biol.* 3:619-625.
- Siden-Kiamos I, Ganter M, Kunze A, Hliscs M, Steinbüchel M, Mendoza J, Sinden RE, Louis C, and Matuschewski K (2011) Stage-specific depletion of myosin A supports an essential role in motility of malarial ookinetes. *Cell Microbiol.* (12):1996-2006.
- Sidhu AB, Verdier-Pinard D, and Fidock DA (2002) Chloroquine resistance in *Plasmodium falciparum* malaria parasites conferred by pfcrt mutations. *Science*. 298(5591):210-3.
- Sidjanski S and Vanderberg JP (1997) Delayed migration of *Plasmodium* sporozoites from the mosquito bite site to the blood. *AmJTrop Med Hyg.* 57:426-429.
- Silvestrini F, Alano P and Williams JL (2000) Commitment to the production of male and female gametocytes in the human malaria parasite *Plasmodium falciparum*. *Parasitology*. 121:465-471.
- Silvestrini F, Bozdech Z, Lanfrancotti A, Di Giulio E, Bultrini E, Picci L, Derisi JL, Pizzi E and Alano P (2005) Genome-wide identification of genes upregulated at the onset of gametocytogenesis in *Plasmodium falciparum*. *Mol Biochem Parasitol.* 143(1):100-10.
- Silvestrini F, Lasonder E, Olivieri A, Camarda G, van Schaijk B, Sanchez M, Younis SY, Sauerwein R and Alano P (2010) Protein Export Marks the Early Phase of Gametocytogenesis of the Human Malaria Parasite *Plasmodium falciparum*. *Mol Cell Proteomics*. 9(7):1437-48.

- Silvie O, Mota M, Matuschewski K, and Prudencio M (2008) Interactions of the malaria parasite and its mammalian host. *Current Opinion in Microbiology* (11):352-359.
- Simon N, Scholz SM, Moreira CK, Templeton TJ, Kuehn A, Dude MA, and Pradel G (2009) Sexual stage adhesion proteins form multi-protein complexes in the malaria parasite *Plasmodium falciparum*. *J Biol Chem*. 284:14537-46.
- Sims JS, Militello KT, Sims PA, Patel VP, Kasper JM, and Wirth DF (2009) Patterns of gene specific and total transcriptional activity during the *Plasmodium falciparum* intraerythrocytic developmental cycle. *Eukaryot Cell*. 8:327-38.
- Sinden RE, and Strong K (1978) An ultrastructural study of the sporogonic development of *Plasmodium falciparum* in *Anopheles gambiae*. *Trans R Soc Trop Med Hyg*. 72(5):477-91.
- Sinden RE (1983) Sexual development of malarial parasites. *Adv. Parasitol.* 22:153-216.
- Sinden RE (1985) A cell biologist's view of host cell recognition and invasion by malarial parasites. *Trans Roy Soc Trop Med Hyg*. 79; 598-605.
- Sinden RE and Hartley RH (1985) Identification of meiotic division of malarial parasites. *J Protozool*. 32; 742-4.
- Sinden RE, Butcher GA, Billker O and Fleck SL (1996) Regulation of infectivity of *Plasmodium* to the mosquito vector. *Adv Parasitol*. 38:54-117.
- Sinden RE and Billingsley PF (2001) *Plasmodium* invasion of mosquito cells: hawk or dove? *TRENDS in Parasitology*. 17:5; 209-211.
- Sinden RE (2009) Malaria, sexual development and transmission: retrospect and prospect. *Parasitology*. 136:1427-1434.
- Sinden RE, Carter R, Drakeley C and Leroy D (2012) The biology of sexual development of *Plasmodium*: the design and implementation of transmission-blocking strategies. *Malaria Journal*. 11:70.
- Sinha A., Hughes KR, Modrzynska KK, Otto TD, Pfander C, Dickens NJ, Religa AA, Bushell E, Graham AL, Cameron R, Kafsack BFC, Williams AE, Llina's M, Berriman M, Billker O, and Waters AP (2014) A cascade of DNA-binding proteins for sexual commitment and development in *Plasmodium*. *Nature*. 507(7491):253-7.
- Sinnis P, and Zavala F (2012) The skin: where malaria infection and the host immune response begin. *Semin Immunopathol* 34:787-792.

- Skarnes WC, Rosen B, West AP, Koutsourakis M, Bushell W, Iyer V, Mujica AO, Thomas M, Harrow J, Cox T, Jackson D, Severin J, Biggs P, Fu J, Nefedov M, de Jong PJ, Stewart AF, and Bradley A (2011) A conditional knockout resource for genome-wide analysis of mouse gene function. *Nature*. 474: 337-342.
- Smalley ME and Brown J (1981) *Plasmodium falciparum* gametocytogenesis stimulated by lymphocytes and serum from infected Gambian children. *Trans R Soc Trop Med Hyg*. 75:316-317.
- Smith TG, Lourenc, OP, Carter R, Walliker D and Ranford-Cartwright LC (2000) Commitment to sexual differentiation in the human malaria parasite, *Plasmodium falciparum*. *Parasitology*. 91: 127-133.
- Smith TG, Walliker D and Ranford-Cartwright LC (2002) Sexual differentiation and sex determination in the apicomplexa. *Trends Parasitol*. 18(7): 315-323.
- Snow RW, Guerra CA, Noor AM, Myint HY and Hay SI (2005) The global distribution of clinical episodes of *Plasmodium falciparum* malaria. *Nature*. 434;214-17.
- Soldati-Favre D (2008) Molecular dissection of host cell invasion by the apicomplexans: the glideosome. *Parasite*. 15:197-205.
- Somsaka V, Srichairatanakool S, Kamchonwongpaisana S, Yuthavongsa Y, and Uthaipibulla C (2011) Small-scale in vitro culture and purification of *Plasmodium berghei* for transfection experiment. *Mol Biochem Parasitol*. 177: 156-159.
- Sonenberg N and Hinnebusch AG (2009) Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell*. 136: 731-745.
- Sorber K, Dimon MT and DeRisi JL (2011) RNA-Seq analysis of splicing in *Plasmodium falciparum* uncovers new splice junctions, alternative splicing and splicing of antisense transcripts. *Nucleic Acids Res*. 39: 3820-3835.
- Sowunmi A, Balogun ST, Gbotosho GO, and Happi CT (2008) *Plasmodium falciparum* gametocyte sex ratios in children with acute, symptomatic, uncomplicated infections treated with amodiaquine. *Malar J*. 2(7):169.
- Steinbuechel M, and Matuschewski K (2009) Role for the *Plasmodium* sporozoite-specific transmembrane protein S6 in parasite motility and efficient malaria transmission. *Cell Microbiol*. 11(2):279-88.
- Sterling CR, Aikawa M and Vanderberg JP (1973) The passage of *Plasmodium berghei* sporozoites through the salivary glands of *Anopheles stephensi*: an electron microscope study. *J Parasitol*. 59: 593-605.

- Stone JR, and Wray GA (2001) Rapid evolution of cis-regulatory sequences via local point mutations. *Mol Biol Evol.* 18:1764-1770.
- Stoute JA, Slaoui M, Heppner DG, Momin P, Kester KE, Desmons P, Wellde BT, Garcon N, Krzych U, and Marchand M (1997) A preliminary evaluation of a recombinant circumsporozoite protein vaccine against *Plasmodium falciparum* malaria. RTS, S malaria vaccine evaluation group. *N Engl J Med.* 336:86-91.
- Strahl BD and Allis CD (2000) The language of covalent histone modifications. *Nature.* 403:41-45.
- Stratton M (2008) Genome resequencing and genetic variation. *Nat Biotechnol.* 26(1):65-6.
- Striepen B, Soldati D, Garcia-Reguet N, Dubremetz JF, and Roos DS (2001) Targeting of soluble proteins to the rhoptries and micronemes in *Toxoplasma gondii*. *Mol Biochem Parasitol.* 113(1):45-53.
- Su X, Hayton K, and Wellems TE (2007) Genetic linkage and association analyses for trait mapping in *Plasmodium falciparum*. *Nature reviews. Genetics.* 8: 497-506.
- Sturm A, Amino R, van de Sand C, Regen T, Retzlaff S, Rennenberg A, Krueger A, Pollok JM, Menard R and Heussler VT (2006) Manipulation of host hepatocytes by the malaria parasite for delivery into liver sinusoids. *Science.* 313:1287-1290.
- Sultan AA, Thathy V, de Koning-Ward TF, and Nussenzweig V (2001) Complementation of *Plasmodium berghei* TRAP knockout parasites using human dihydrofolate reductase gene as a selectable marker. *Mol Biochem Parasitol.* 113(1):151-6.
- Swain MT, Tsai IJ, Assefa SA, Newbold C, Berriman M, and Otto TD (2012) A post-assembly genome-improvement toolkit (PAGIT) to obtain annotated genomes from contigs. *Nat Protoc.* 7(7):1260-84.
- Talman AM, Domarle O, McKenzie FE, Ariey F and Robert V (2004) Gametocytogenesis: the puberty of *Plasmodium falciparum*. *Malar. J.* 3:24.
- Talman AM, Paul REL, Sokhna CS, Domarle O, Ariey F, Trape J-F and Robert V (2004a) Influence of chemotherapy on *Plasmodium* gametocytes sex ratio of Mice and Men. *Am J Trop Med Hyg.* 71(6);739-744.
- Tanay A (2006) Extensive low-affinity transcriptional interactions in the yeast genome. *Genome Res.* 16, 962-972.

- Tanner M and de Savigny D (2008) Malaria eradication back on the table. *Bull World Health Organ.* 86:2;82.
- Targett GA, Harte PG, Eida S, Rogers NC and Ong CS (1990) *Plasmodium falciparum* sexual stage antigens: immunogenicity and cell-mediated responses. *Immunol Lett.* 25:77-81.
- Tarun AS, Peng X, Dumpit RF, Ogata Y, Silva-Rivera H, Camargo N, Daly TM, Bergman LW, and Kappe SH (2008) A combined transcriptome and proteome survey of malaria parasite liver stages. *Proc Natl Acad Sci USA.* 105:305-10.
- Taylor CJ, McRobert L, and Baker DA (2008) Disruption of a *Plasmodium falciparum* cyclic nucleotide phosphodiesterase gene causes aberrant gametogenesis, *Mol. Microbiol.* 69:110-118.
- Taylor LH, Walliker D and Read AF (1997) Mixed-genotype infections of the rodent malaria *Plasmodium chabaudi* are more infectious to mosquitoes than single-genotype infections. *Parasitology.* 115:121-132.
- Teboh-Ewungkem MI, and Wang M (2012) Male fecundity and optimal gametocyte sex ratios for *Plasmodium falciparum* during incomplete fertilization. *J Theor Biol.* 307:183-92.
- Teif VB, and Rippe K (2011) Nucleosome mediated crosstalk between transcription factors at eukaryotic enhancers. *Phys Biol.* 8(4):044001.
- Templeton TJ, Keister DB, Muratova O, Procter JL, and Kaslow DC (1998) Adherence of erythrocytes during exflagellation of *Plasmodium falciparum* microgametes is dependent on erythrocyte surface sialic acid and glycophorins. *J Exp Med.* 187(10):1599-609.
- Templeton TJ, Iyer LM, Anantharaman V, Enomoto S, Abrahante JE, Subramanian GM, Hoffman SL, Abrahamsen MS and Aravind L (2004) Comparative analysis of *apicomplexa* and genomic diversity in eukaryotes. *Genome Res.* 14: 1686-1695.
- Tewari R, Dorin D, Moon R, Doerig C, and Billker O (2005) An atypical mitogen-activated protein kinase controls cytokinesis and flagellar motility during male gamete formation in a malaria parasite. *Mol Microbiol.* 58(5): 1253-63.
- Tewari R, Straschil U, Bateman A, Böhme U, Cherevach I, Gong P, Pain A, and Billker O (2010) The systematic functional analysis of *Plasmodium* protein kinases identifies essential regulators of mosquito transmission. *Cell Host Microbe.* 8(4):377-87.

- Thathy V, and Menard R (2002) Gene targeting in *Plasmodium berghei*. *Methods Mol. Med.* 72: 317-331.
- The malERA Consultative Group on Basic Science and Enabling Technologies (2011) A Research Agenda for Malaria Eradication: Basic Science and Enabling Technologies. *PLoS Med* 8(1): e1000399.
- Thompson J, and Sinden RE (1994) In situ detection of Pbs21 mRNA during sexual development of *Plasmodium berghei*. *Mol Biochem Parasitol.* 68(2):189-96.
- Thompson J, Janse CJ and Waters AP (2001) Comparative genomics in *Plasmodium*: a tool for the identification of genes and functional analysis. *Mol. Biochem Parasitol.* 118;147-154.
- Thompson J, Fernandez-Reyes D, Sharling L, Moore SG, Eling WM, Kyes SA, Newbold CI, Kafatos FC, Janse CJ, and Waters AP (2007) *Plasmodium* cysteine repeat modular proteins 1-4: complex proteins with roles throughout the malaria parasite life cycle. *Cell Microbiol* 9: 1466-1480.
- Tilley L and McConville M (2013) Sensing when it's time for sex. *Nature* (499) 38.
- Tomas AM, Margos G, Dimopoulos G, van Lin LHM, de Koning-Ward TF, Sinha R, Lupetti P, Beetsma AL, Rodriguez MC, Karras M, Hager A, Mendoza J, Butcher GA, Kafatos F, Janse CJ, Waters AP, and Sinden RE (2001) P25 and P28 proteins of the malaria ookinete surface have multiple and partially redundant functions. *EMBO Journal.* 20:15;3975-3983.
- Tonkin CJ, Carret CK, Duraisingh MT, Voss TS, Ralph SA, Hommel M, Duffy MF, Silva LM, Scherf A, Ivens A, Speed TP, Beeson JG, and Cowman AF (2009) Sir2 paralogues cooperate to regulate virulence genes and antigenic variation in *Plasmodium falciparum*. *PLoS Biol* 7: e1000084.
- Toyé PJ, Sinden RE, and Canning EU (1977) The action of metabolic inhibitors on microgametogenesis in *Plasmodium yoelii nigeriensis*. *Z Parasitenkd.* 53(2):133-41.
- Trager W and Jensen JB (1976) Human malaria parasite in continuous culture. *Science.* 193;673-675.
- Trager W (1979) *P. falciparum* in culture: improved continuous flow method. *J. Protozool.* 26(1): 25-29.
- Trager W and Gill GS (1989) *Plasmodium falciparum* gametocyte formation in vitro: its stimulation by phorbol diesters and by 8- bromo cyclic adenosine monophosphate. *J Protozool.* 36:451-454.

- Trager W and Gill GS (1992) Enhanced gametocyte formation in young erythrocytes by *Plasmodium falciparum* in vitro. *J Protozool.* 39:429-432.
- Trager W, Gill GS, Lawrence C and Nagel RL (1999) *Plasmodium falciparum*: enhanced gametocyte formation in vitro in reticulocyte-rich blood. *Exp Parasitol.* 91:115-118.
- Trelle MB, Salcedo-Amaya AM, Cohen AM, Stunnenberg HG, and Jensen ON (2009) Global histone analysis by mass spectrometry reveals a high content of acetylated lysine residues in the malaria parasite *Plasmodium falciparum*. *Journal of Proteome Research.* 8: 3439-3450.
- Trueman HE, Raine JD, Florens L, Dessens JT, Mendoza J, Johnson J, Waller CC, Delrieu I, Holders AA, Langhorne J, Carucci DJ, Yates JR, and Sinden RE (2004) Functional characterization of an LCCL-lectin domain containing protein family in *Plasmodium berghei*. *J Parasitol.* 90(5):1062-71.
- Tsai IJ, Otto TD, and Berriman M (2010) Improving draft assemblies by iterative mapping and assembly of short reads to eliminate gaps. *Genome Biol.* 11(4):R41.
- Tucker T, Marra M, and Friedman JM (2009) Massively Parallel Sequencing: The Next Big Thing in Genetic Medicine. *The American Journal of Human Genetics.* 85:142-154.
- Tuteja R, Ansari A, and Chauhan VS (2011) Emerging functions of transcription factors in malaria parasite. *J Biomed Biotechnol.* 461979.
- Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, Pertz LM, Haugen E, Hayden H, Albertson D, Pinkel D, Olson MV, and Eichler EE (2005) Fine-scale structural variation of the human genome. *Nat Genet.* 37:727-732.
- Upadhyay R, Bawankar P, Malhotra D and Patankar S (2005) A screen for conserved sequences with biased base composition identifies noncoding RNAs in the A-T rich genome of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* 144: 149-158.
- Vaidya AB, Muratova O, Guinet F, Keister D, Wellems TE and Kaslow DC (1995) A genetic locus of *Plasmodium falciparum* chromosome 12 linked to a defect in mosquito-infectivity and male gametogenesis. *Mol. Biochem. Parasitol.* 69:65-71.
- Vakoc CR, Letting DL, Gheldof N, Sawado T, Bender MA, Groudine M, Weiss MJ, Dekker J, and Blobel GA (2005) Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Mol. Cell.* 17: 453-462.



- van der Kolk M, Tebo AE, Nimpaye H, Ndombol DN, Sauerwein RW, and Eling WM (2003) Transmission of *Plasmodium falciparum* in urban Yaoundé, Cameroon, is seasonal and age-dependent. *Trans R Soc Trop Med Hyg.* 97(4):375-9.
- van Dijk MR, Waters AP, and Janse CJ (1995) Stable transfection of malaria parasite blood stages. *Science.* 268(5215):1358-62.
- van Dijk MR, Janse CJ, and Waters AP (1996) Expression of a *Plasmodium* gene introduced into subtelomeric regions of *Plasmodium berghei* chromosomes. *Science.* 271: 662-665.
- van Dijk MR, Janse CJ, Thompson J, Waters AP, Braks JAM, Dodemont HJ, Stunnenberg HG, van Gemert GJ, Sauerwein RW, and Eling W (2001) A Central Role for P48/45 in Malaria parasite male gamete fertility *Cell.* 104;153-164.
- van Dijk MR, van Schaijk BCL, Khan SM, van Dooren MW, Ramesar J, Kaczanowski S, van Gemert GJ, Kroeze H, Stunnenberg HG, Eling WM, Sauerwein RW, Waters AP, and Janse CJ (2010) Three Members of the 6-cys Protein Family of *Plasmodium* Play a Role in Gamete Fertility. *PLoS Pathog* 6(4): e1000853.
- van Ooij C (2008) Hidden treasure uncovered? *Nature Reviews Microbiology.* 6: 570.
- van Ooij C (2011) Parasitology: Basigin opens the door to malaria. *Nat Rev Microbiol.* 10(1):3.
- Vanderberg JP, and Frevert U (2004) Intravital microscopy demonstrating antibody-mediated immobilisation of *Plasmodium berghei* sporozoites injected into skin by mosquitoes. *Int J Parasitol.* 34(9):991-6.
- Vaughan AM, Aly ASI and Kappe SHI (2008) Malaria parasite pre-erythrocytic stage infection: Gliding and Hiding. *Cell Host Microbe.* 4(3):209-218.
- Vincke IH and Lips M (1948) Un nouveau *Plasmodium* d'un rongeur sauvage du Congo: *Plasmodium berghei* n.sp. *Annales de la Société Belge de Médecine Tropicale.* 28; 97-104.
- Vlachou D, Lycett G, Sidén-Kiamos I, Blass C, Sinden RE, and Louis C (2001) *Anopheles gambiae* laminin interacts with the P25 surface protein of *Plasmodium berghei* ookinetes. *Mol Biochem Parasitol.* 112(2):229-37.
- Vlachou D, Zimmermann T, Cantera R, Janse CJ, Waters AP, and Kafatos FC (2004) Real-time, in vivo analysis of malaria ookinete locomotion and mosquito midgut invasion. *Cell. Microbiol.* 6;671-685.

- Volz JC, Bartfai R, Petter M, Langer C, Josling GA, Tsuboi T, Schwach F, Baum J, Rayner JC, Stunnenberg HG, Duffy MF, and Cowman AF (2012) PfSET10, a *Plasmodium falciparum* methyltransferase, maintains the active var gene in a poised state during parasite division. *Cell Host Microbe*. 11: 7-18.
- Voss TS, Kaestli M, Vogel D, Bopp S, and Beck HP (2003) Identification of nuclear proteins that interact differentially with *Plasmodium falciparum* var gene promoters. *Mol Microbiol*. 48:1593-1607.
- Voss TS, Healer J, Marty AJ, Duffy MF, Thompson JK, Beeson JG, Reeder JC, Crabb BS, and Cowman AF (2006) A var gene promoter controls allelic exclusion of virulence genes in *Plasmodium falciparum* malaria. *Nature*. 439: 1004-1008.
- Voza T, Miller JL, Kappe SH, and Sinnis P (2012) Extrahepatic exoerythrocytic forms of rodent malaria parasites at the site of inoculation: clearance after immunization, susceptibility to primaquine and contribution to blood stage infection. *Infect Immun*. 80:2158-2164.
- Wang J, Sarov M, Rientjes J, Fu J, Hollak H, Kranz H, Xie W, Stewart AF, and Zhang Y (2006) An improved recombineering approach by adding RecA to lambda Red recombination. *Mol. Biotechnol*. 32: 43-53.
- Warren RL, Sutton GG, Jones SJ and Holt RA (2007) Assembling millions of short DNA sequences using SSAKE. *Bioinformatics*. 23:500-501.
- Waters AP, Syin C and McCutchan TF (1989) Developmental regulation of stage-specific ribosome populations in *Plasmodium*. *Nature*. 342:438-40.
- Waters AP, van Spaendonk RM, Ramesar J, Vervenne RA, Dirks RW, Thompson J and Janse CJ (1997) Species-specific regulation and switching of transcription between stage-specific ribosomal RNA genes in *Plasmodium berghei*. *J. Biol. Chem*. 272:3583-3589.
- Weatherall DJ, Miller LH, Baruch DI, Marsh K, Doumbo OK, Casals- Pascual, C and Roberts DJ (2002) Malaria and the red cell. *Hematology Am. Soc. Hematol. Educ. Program*. 1:35-57.
- Wellems TE, Hayton K and Fairhurst RM (2009) The impact of malaria parasitism: from corpuscles to communities. *J. Clin. Invest*. 119:2496-2505.
- Whiteford N, Haslam N, Weber G, Prügel-Bennett A, Essex JW, Roach PL, Bradley M and Neylon C (2005) An analysis of the feasibility of short read sequencing. *Nucleic Acids Res*. 33:e171.

- Wickens M, Bernstein DS, Kimble J, and Parker R (2002) A PUF family portrait: 3'UTR regulation as a way of life. *Trends Genet* 18: 150-157.
- Williams JL (1999) Stimulation of *Plasmodium falciparum* gametocytogenesis by conditioned medium from parasite cultures. *Am J Trop Med Hyg.* 60:7-13.
- Winograd E, Clavijo CA, Bustamante LY and Jaramillo M (1999) Release of merozoites from *Plasmodium falciparum* infected erythrocytes could be mediated by a non-explosive event. *Parasitol. Res.* (85); 621-624.
- Wold B and Myres RM (2008) Sequence census methods for functional genomics. *Nature Methods.* 5(1): 19-21.
- World Health organization, WHO (2000) Severe *falciparum* malaria. *Trans R Soc Trop Med Hyg.* 94: S1/1-S1/90.
- World Health Organization (2013) Immunization, Vaccines and Biologicals. Internet website  
[http://www.who.int/vaccine\\_research/Malaria/en/index.html](http://www.who.int/vaccine_research/Malaria/en/index.html)
- World Malaria Report (2009) World Health Organization. WHO Library Cataloguing-in-Publication Data, World Health Organization, Geneva, Switzerland. ISBN 978 92 4 156390 1.
- World Malaria Report (2012) World Health Organization. WHO Library Cataloguing-in-Publication Data, World Health Organization, Geneva, Switzerland. ISBN 978 92 4 156453 3
- World Malaria Report (2013) World Health Organization. WHO Library Cataloguing-in-Publication Data, World Health Organization, Geneva, Switzerland. ISBN 978 92 4 156469 4
- Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV and Romano LA (2003) The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol.* 20(9):1377-1419.
- Wray GA (2007) The evolutionary significance of cis-regulatory mutations. *Nature Reviews Genetics.* 8: 206-216.
- Wu Y, Kirkman LA, and Wellems TE (1996) Transformation of *Plasmodium falciparum* malaria parasites by homologous integration of plasmids that confer resistance to pyrimethamine. *Proc. Natl. Acad. Sci.* 93: 1130-1134.
- Yamauchi LM, Coppi A, Snounou G and Sinnis P (2007) *Plasmodium* sporozoites trickle out of the injection site. *Cell Microbiol.* 9:1215-1222.

- Yang A, Zhu Z, Kapranov P, McKeon F, Church GM, Gingeras TR, and Struhl K (2006) Relationships between p63 binding, DNA sequence, transcription activity, and biological function in human cells. *Mol Cell*. 24: 593-602.
- Yekutieli P (1980) The global malaria eradication campaign In: *Eradication of infectious diseases: a critical study*. Contributions to epidemiology and biostatistics. Basel, Switzerland: Karger. pp. 34-88.
- Yeoh S, O'Donnell RA, Koussis K, Dluzewski AR, Ansell KH, Osborne SA, Hackett F, Withers-Martinez C, Mitchell GH, and Bannister LH (2007) Subcellular discharge of a serine protease mediates release of invasive malaria parasites from host erythrocytes. *Cell*. 131:1072-1083.
- Yochum GS, Sherrick CM, Macpartlin M, and Goodman RH (2010) A beta-catenin/TCF-coordinated chromatin loop at MYC integrates 5' and 3' Wnt responsive enhancers. *Proc Natl Acad Sci U S A*. 107(1):145-50.
- Yoon S, Xuan Z, Makarov V, Ye K and Sebat J (2009) Sensitive and accurate detection of copy number variants using read depth of coverage. *Genome Res*. 19(9):1586-92.
- Young JA, Fivelman QL, Blair PL, de la Vega P, Le Roch KG, Zhou Y, Carucci DJ, Baker DA, and Winzeler EA (2005) The *Plasmodium falciparum* sexual development transcriptome: a microarray analysis using ontology-based pattern identification. *Mol Biochem Parasitol*. 143: 67-79.
- Yuda M, Sawai T and Chinzei Y (1999) Structure and Expression of an Adhesive Protein-like Molecule of Mosquito Invasive-stage Malarial Parasite. *J. Exp. Med*. 189:1947-1952.
- Yuda M, Iwanaga S, Shigenobu S, Mair GR, Janse CJ, Waters AP, Kato T, and Kaneko I (2009) Identification of a transcription factor in the mosquito-invasive stage of malaria parasites. *Mol Microbiol*. 71:1402-14.
- Yuda M, Iwanaga S, Shigenobu S, Kato T, and Kaneko I (2010) Transcription factor AP2-Sp and its target genes in malarial sporozoites. *Mol Microbiol*. 75(4):854-63.
- Zamore PD, Williamson JR, and Lehmann R (1997) The Pumilio protein binds RNA through a conserved domain that defines a new class of RNA-binding proteins. *RNA* 3: 1421-1433.
- Zaret K (2005) Chromatin Assembly and Analysis - Micrococcal Nuclease Analysis of Chromatin Structure. *Current Protocols in Molecular Biology*. Supplement 69: 21.1.1-21.1.17.

- Zaret K, and Carroll J (2011) Pioneer transcription factors: establishing competence for gene expression *Genes Dev.* 25: 2227-2241.
- Zerbino DR and Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18:821-9.
- Zhang B, Gallegos M, Puoti A, Durkin E, Fields S, Kimble J, and Wickens MP (1997) A conserved RNA-binding protein that regulates sexual fates in the *C. elegans* hermaphrodite germ line. *Nature.* 390: 477-484.
- Zhang Y, Buchholz F, Muirers JP and Stewart AF (1998) A new logic for DNA engineering using recombination in *Escherichia coli*. *Nat. Genet.* 20: 123-128.
- Zhang M, Fennell C, Ranford-Cartwright L, Sakthivel R, Gueirard P, Meister S, Caspi A, Doerig C, Nussenzweig RS, Tuteja R, Sullivan WJ Jr, Roos DS, Fontoura BM, Ménard R, Winzeler EA, and Nussenzweig V (2010) The *Plasmodium* eukaryotic initiation factor-2alpha kinase IK2 controls the latency of sporozoites in the mosquito salivary glands. *J Exp Med.* 207(7):1465-74.
- Zhang X, Tolzmann CA, Melcher M, Haas BJ, Gardner MJ, Smith JD and Feagin JE (2011) Branch point identification and sequence requirements for Intron splicing in *Plasmodium falciparum*. *Eukaryotic Cell.* 10(11):1422.
- Zhang Y, Sicot G, Cui X, Vogel M, Wuertzer CA, Lezon-Geyda K, Wheeler J, Harki DA, Muzikar KA, Stolper DA, Dervan PB, and Perkins AS (2011) Targeting a DNA binding motif of the EVI1 protein by a pyrrole-imidazole polyamide. *Biochemistry.* 50(48):10431-41.
- Zhou H, Hu H, and Lai M (2010) Non-coding RNAs and their epigenetic regulatory mechanisms. *Biol Cell.* 102: 645-655.
- Zou J, Liu X, Shi T, Huang X, Wang H, Hao L, Yin G and Suo X (2009) Transfection of *Eimeria* and *Toxoplasma* using heterologous regulatory sequences. *Int. J. Parasitol.* 39: 1189-1193.